A Notion of Prominence for Games with Natural-Language Labels

Alessandro Sontuoso *and* Sudeep Bhatia

September 20, 2018

Abstract. We study games with natural-language labels (i.e., strategic problems where options are denoted by words), for which we propose and test a measurable characterization of prominence. We assume that – ceteris paribus – players find particularly prominent those strategies that are denoted by labels frequently used in everyday language: to operationalize this assumption, we suggest that the prominence of a strategy-label is correlated with its frequency of occurrence in large text corpora. In order to test for the strategic use of word frequency, we consider experimental games with different incentive structures (such as incentives *to* and *not to* coordinate), as well as subjects from different cultural/linguistic backgrounds. We find that frequently-mentioned labels are more (less) likely to be selected when there are incentives to match (mismatch) others. Furthermore, varying one's knowledge of the others' cultural background significantly affects one's reliance on word frequency. Overall, our studies suggest that individuals select strategies that fulfill our characterization of prominence in a (boundedly) rational manner.

KEYWORDS: focal points, salience, coordination, hide-and-seek, level-*k*.

JEL Classification Numbers: C72, C91.

Contact: Alessandro Sontuoso, *Philosophy, Politics and Economics*, University of Pennsylvania, 249 S. 36th St., Philadelphia, Pennsylvania, 19104, sontuoso@sas.upenn.edu; Sudeep Bhatia, *Department of Psychology*, University of Pennsylvania, 425 S. University Ave., Philadelphia, Pennsylvania, 19104, bhatiasu@sas.upenn.edu.

I - Introduction

Coordination problems are known to pervade social and economic interactions, yet are often solved by virtue of shared "cultural" understandings (Camerer and Knez, 1997; Weber and Camerer, 2003). For example, global companies that wish to match potential demand in developing markets have to adjust their strategies in such a way to connect to (and hence coordinate with) culturally-diverse consumers. In that case, cultural awareness substantially affects an organization's marketing and customer support strategy – from labelling to customer services – thereby making *some* products more salient for *some* consumer segments (Kapferer, 2012).

Thomas Schelling (1960) was the first to note that a wide range of everyday interactions can formally be represented as a coordination problem (i.e., a symmetric, simultaneous-move game with multiple pure-strategy Nash equilibria).[1] Schelling observed that problems of this kind are generally solved by resorting to either *contextual* or *cultural* cues, which drive expectations in such a way to make a particular course of action salient for a particular group of people. Here, focusing on games with natural-language labels (i.e., strategic problems where options are denoted by words), we propose and test a characterization of prominence that explicitly rests on players' culture.

Since Schelling's initial work, the study of prominence (i.e., salience) has generally focused on the strategic use of contextual – but culture-invariant – cues. That is, cues that drive one's perceptions as to the uniqueness of a solution, regardless of one's culture. For example, Harsanyi and Selten's (1988) payoff-dominance criterion assumes that *any* rational player (upon facing a one-shot game with no pre-play communication) would discard solutions that are collectively suboptimal: in that case, the payoff structure of the game serves as a "cue", thereby directing players' expectations toward the collectively optimal solution. Similar characterizations of prominence incorporate the payoff-dominance criterion into a broader theory of framing, whereby players select the collectively optimal solution to some private description of the game

[1] The class of coordination problems contains any situation in which there are multiple ways agents may "match" their actions for mutual benefit. Such a class contains a vast and diverse array of interactions, including games *with* Pareto-rankable equilibria (e.g., the Hi-Lo game; Bacharach, 2006) as well as *without* Pareto-rankable equilibria (e.g., the driving game; Schelling, 1960). For an early book-length account of coordination games, see Lewis (1969). For a discussion of Schelling's ideas see Sugden and Zamarrón (2006); for an early test of those ideas see Mehta, Starmer, and Sugden (1994).

(Crawford and Haller, 1990; Bacharach, 1993; Sugden, 1995; Casajus, 2000; Blume, 2000; Bacharach and Stahl, 2000; Janssen, 2001; Alós-Ferrer and Kuzmics, 2013).[2] In this regard, lab experiments have confirmed that *frame* and *payoff* asymmetries do (each to a different extent) influence subjects' perception of the game, thereby making some options more prominent than others. For example, think of a coordination game where one of the solutions is an "oddity", in the sense that its label profile or the associated payoff profile differs from the other profiles (Bacharach and Bernasconi, 1997; Rubinstein, Tversky, and Heller, 1997; Crawford, Gneezy, and Rottenstreich, 2008). There the oddity serves as a cue, thus facilitating players' successful coordination.

We stress that according to the above explanations of prominence, a solution is perceived as a "focal point" by virtue of a (contextual) cue that is typically culture-invariant. In what follows, on the other hand, we specify and test a characterization of prominence that depends on individuals' culture. That is, we assume that – ceteris paribus – individuals find particularly prominent those strategies that are denoted by words frequently used in everyday language. To operationalize this assumption, we suggest that the prominence of a strategy-label is correlated with its frequency of occurrence in large text corpora (such as the Google Books corpus). In order to test for the strategic use of word frequency, we consider experimental games with different incentive structures (such as incentives *to* and *not to* coordinate) as well as subjects from different cultural/linguistic backgrounds. Such a characterization informs a rich and previously untested set of hypotheses.

Before elaborating on our experiments, we stress that this paper departs from the studies mentioned above in a few important ways. First, we consider games with no obvious *frame* asymmetries:[3] by focusing on problems where each strategy is denoted by a label relating to the same category/frame (e.g., names of food ingredients), we are able to rule out a common driver of prominence as an explanation for our data patterns. Furthermore, we focus on games with no

[2] Theoretical accounts of coordination are sometimes divided into two broad families, namely, team reasoning and level-*k* theories. Models of team reasoning rest on the assumption that every player follows the decision rule which, if followed by all other players, would be optimal for each of them (e.g., Sugden, 1995). On the other hand, level-*k* theories assume a hierarchy of cognitive levels, whereby higher types best respond to the actions of lower-level players (anchoring their beliefs in the behavior of strategically naïve individuals; e.g., for a level-*k* model of coordination, see Bacharach and Stahl, 2000). For discussion, see Bardsley, Mehta, Starmer, and Sugden (2010) and Crawford, Costa-Gomes, and Iriberri (2013).

[3] For recent coordination experiments involving *more or less* obvious frame asymmetries, see: Blume and Gneezy (2010); Bardsley et al. (2010); Hargreaves Heap, Rojo Arjona, and Sugden (2014).

payoff asymmetries;[4] more specifically, we study one-shot games without Pareto-rankable equilibria, thereby ruling out another common driver of prominence. For example, consider a "pure" coordination game where strategies are denoted by the labels {paprika, curry, chili}.[5] Given this, we aim to address questions that have remained unanswered by previous research, such as the following: Is it possible to define an a-priori *measurable* proxy for prominence in games with no obvious odd-one-outs? Does one's culture affect such a proxy, and hence one's choice behavior, under different incentive structures? Further, does one's knowledge of the others' cultural background affect one's own choice behavior?

In short, for the first time we use the frequency of occurrence of a word in the Google Books corpus ("n-gram frequency") as a proxy for the prominence of a strategy-label.[6] We note that Google Books n-gram frequency is a standard metric for word popularity in natural languages, and is widely used for linguistic, psychological, sociological, and historical research (Michel et al., 2011). Our approach is founded on a psychologically grounded characterization of prominence, that has practical implications for our understanding of strategic reasoning. Indeed, research from cognitive psychology has shown that the *frequency of exposure* to words is closely related to word *fluency*, that is, the ease with which an individual is able to recognize, retrieve, and process a word. Word frequency – through its effect on fluency – has been shown to have an important role in a wide range of memory and language tasks (e.g., Anderson and Schooler, 1991; Balota and Spieler, 1999; Jescheniak and Levelt, 1994; Morrison and Ellis, 1995; Seidenberg and McClelland, 1989). In particular, there is evidence that individuals use word

[4] For early experiments on coordination games with Pareto-rankable equilibria, see Cooper, DeJong, Forsythe, and Ross (1990); repeated games are studied in Van Huyck, Battalio, and Beil (1990). For recent experiments involving Pareto-rankable equilibria, see Bardsley and Ule (2017) and Faillo, Smerilli, and Sugden (2017). A related line of research investigates focal points in the context of bargaining problems: see for example Isoni, Poulsen, Sugden, and Tsutsui (2013).

[5] Pure coordination games present the following payoff structure: if players select the same strategy-label, they each receive an identical – positive – payoff (say, 1 currency unit); otherwise, they each receive nothing.

[6] Sugden (1995) was the first to explicitly suggest that players may distinguish among labelled strategies in terms of the frequency with which they have heard them (pp. 547-548), but did not systematically test this conjecture. In this regard, Hargreaves Heap, Rojo Arjona, and Sugden's (2017) experimental study of coordination assumes that players base their choice decisions on one of several *selection rules* (e.g., "choose your favorite", "choose the odd-one-out", "choose the most frequently mentioned", etc.). In order for the experimenter to estimate which strategy-label is dictated by each rule, Hargreaves Heap et al. have a separate group of subjects answer a questionnaire (e.g., "Which option do you think would be selected by this rule?"). While related, we note that their study does not address the questions we posed above, as their design does not vary subjects' cultural background *or* subjects' knowledge of the others' background. We also note that Hargreaves Heap et al.'s ex-post-facto estimation of the selection rules does not generate an a-priori measurable proxy for prominence.

frequency as a cue, in several *non-strategic* judgment and decision-making domains such as probability judgment (Dougherty, Franco-Watkins, and Thomas, 2008; Tversky and Kahneman, 1974), risk perception (Hertwig, Pachur, and Kurzenhäuser, 2005; Lichtenstein, Slovic, Fischhoff, Layman, and Combs, 1978), and factual judgment (Gigerenzer and Goldstein, 1996; Hertwig, Herzog, Schooler, and Reimer, 2008). In all these domains, the frequency of occurrence of a word in everyday language is positively correlated with the tendency to select that word as a response, and to evaluate the object denoted by that word as being large, important, truthful, or desirable. In fact, fluency is thought to be one of the mechanisms through which the *availability heuristic* operates (Schwarz, Bless, Strack, Klumpp, Rittenauer-Schatka, and Simons, 1991; Tversky and Kahneman, 1973).

The findings above suggest that word frequency may be useful in strategic domains as well. For example, consider the aforementioned coordination game with strategy set {paprika, curry, chili}. There, the option that is more frequently mentioned in everyday language may be more "fluent", that is, easy to process: so we might expect that individuals will be drawn to such option, the same way as they are drawn to fluent options in non-strategic problems. Most importantly, unlike in the decision-making literature, our goal here is to test whether subjects *strategically* exploit word frequency. One mechanism that supports our hypothesis (about the strategic use of word frequency) is that individuals may realize that their matched participants might be culturally alike, and hence view the problem in a similar way (Hedden and Zhang, 2002; Sebanz, Knoblich, and Prinz, 2003; Rubinstein and Salant, 2016). Thus, if there is common reason to believe that an option easily comes to mind to people with the same cultural background, then it might be optimal to select that option in coordination games (Lewis, 1969; Morris and Shin, 1997; Cubitt and Sugden, 2003). In order to test for the strategic use of word frequency, below we consider games with different incentive structures as well as subjects from different cultural/linguistic backgrounds. If subjects make use of word frequency for strategic reasons, then we should expect that varying incentives may affect choice behavior; similarly, we should expect that varying one's knowledge of the others' cultural background may affect one's choice behavior.

We designed our Study 1 as a preliminary test for predicting behavior in 2-player pure coordination games with a finite set of options (such games are sometimes referred to as "matching games"). Having found a strong correlation between word frequency and choice behavior, we designed two more studies to explicitly test for the strategic use of word frequency.

In particular, Study 2 contrasts choice behavior in *(i)* a pure coordination game against the behavior of participants in three alternative roles/conditions (with each condition featuring exactly the same list of labels). Specifically, we considered: *(ii)* the case in which a subject is prompted to *pick* an option, without the explicit objective to match someone else's choice (i.e., a non-strategic problem); *(iii)* the case in which a subject is prompted to *avoid* matching her counterpart's choice, under the assumption that her counterpart instead wants to match (i.e., the role of "Hider" in a Hide-and-Seek game); *(iv)* the case in which a subject is prompted to *match* her counterpart's choice, under the assumption that her counterpart instead wants to avoid any such match (i.e., the role of "Seeker" in a Hide-and-Seek game). Since all our conditions involved the exact same options, if the effect of word frequency in Study 1 were merely due to an automatic (or naïve) response, then we should observe similar choice distributions across the four conditions. In short, our data show that participants in problem *ii* (i.e., "Pickers") were about as likely to select the most frequently-mentioned label as were participants in the coordination game ("Coordinators"). By contrast, Hiders were less likely to select the most frequently-mentioned label than Seekers and, in turn, Seekers were less likely than Coordinators. This pattern suggests a boundedly rational use of strategy-labels that is consistent with a particular specification of level-*k* reasoning (Nagel, 1995; Stahl and Wilson, 1995; Costa-Gomes, Crawford, and Broseta, 2001; Camerer, Ho, and Chong, 2004; Crawford and Iriberri, 2007).

Finally, we designed our third study to delve further into the strategic use of word frequency in coordination games. To that end, we varied the cultural/linguistic makeup of our subject pool; that is, we recruited subjects residing in either the US or the UK, and we manipulated their knowledge of the counterpart's country of residence. Specifically, we had participants play a set of (one-shot) coordination games, with each game involving three strategy-labels that we purposely selected in such a way that the option with the highest word frequency varied between the American- and British-English vocabularies (as measured by the labels' NGRAM in the American- and British-English Google Books corpora, respectively).[7]

[7] For example, consider the aforementioned coordination game with strategy set {paprika, curry, chili}. There, "curry" has the highest NGRAM in *British English* but the lowest one in American English; conversely, "chili" has the highest NGRAM in *American English* and the lowest one in British English.

Consistent with our predictions, the data show that *choice behavior differed between US and UK subjects*, and in each case it was positively related to the word frequency of the strategy-labels in the relevant vocabulary. Also, the data show that subjects were less likely to rely on word frequency as a means to guiding their behavior whenever they knew that their counterpart resided in a different country. Put differently, when subjects knew that their assigned partner was alike (in terms of cultural background), they were substantially more likely to select the label most frequently mentioned in their own vocabulary! We further compared coordination rates that would be obtained if different subsamples were paired with each other, using Monte Carlo methods: in a nutshell, successful coordination was more likely when participants were *knowingly* paired with partners from their own country, as opposed to when they were knowingly or unknowingly paired with partners from a different country. In particular, we found that subjects who were knowingly paired with partners from the same country exhibited expected coordination rates between 10 and 20 percentage points higher than chance.

To conclude, for the first time we study the impact of word frequency on strategic play, controlling for subjects' cultural background. Our experiments provide very robust evidence indicating that individuals select strategies fulfilling our notion of prominence, and they furthermore do so in a (boundedly) rational manner. Most notably – in the case of coordination games – reliance on word frequency leads to higher rates of coordination than chance, and more so when individuals knowingly share a cultural background. The remainder of the article is organized in this manner: section II lays out the experimental procedures, sections III-V present our studies, and section VI concludes.

II - General Procedures

Our studies were conducted online between September and November 2016. We note that a primary motivation for running online experiments is the ease with which the experimenter can control and vary the cultural makeup of the subject pool, along with participants' knowledge of the subject pool's cultural characteristics. Such features make online experimentation optimal for testing culture-related hypotheses. (For a methodological discussion of extra-laboratory experiments, we refer the reader to Charness, Gneezy, and Kuhn, 2013.) Specifically, our participants were recruited through the Oxford-based crowdsourcing platform *Prolific Academic.* Participation in our study was limited to individuals with a Prolific-Academic approval rate greater than 95%. At the beginning of each study, subjects were informed that every participant would be assigned a partner at random, and that they would not know the identity of their partner or be able to communicate with her. No subject was allowed to participate in more than one study. Participants' responses were incentivized, as specified in the sections below. (For the experimental instructions and screen shots, see Appendix B.)

As a proxy for the prominence of strategy-labels, we used the (case-insensitive) average yearly "n-gram frequency" of the corresponding word in the Google Books corpus, for books published after 2000. Such NGRAM metrics were obtained through the Google n-gram tool (https://books.google.com/ngrams/) in August 2016, shortly before running the studies. In Studies 1 and 2 we used the general English corpus, whereas in Study 3 we used the American-English and British-English corpora. We stress that Google Books n-gram frequency is a reliable, standard metric for word popularity in corpus linguistics (Michel et al., 2011).[8]

III - Study 1

Demographics. The subject pool for Study 1 consisted of 91 US resident individuals. The average participant was 33 years old, and 57% of the subjects were male. Participants took less than 10 minutes to review the instructions and complete all the tasks. Subjects received a 0.5 GBP participation fee (in addition to the payoffs earned in each game), which is on par with typical wages on Prolific Academic or other Internet marketplaces such as MTurk.

Design. We designed Study 1 as a preliminary test for predicting behavior in 2-player pure coordination games. Specifically, the study involved a series of (one-shot) coordination games, with each game featuring a 3-element strategy set such that: each member of a pair received GBP 0.10 if both players chose the same option; each member of a pair received nothing otherwise. Figure 1 represents the game structure in bimatrix form (there, for expositional purposes, the set of strategies is denoted by {X, Y, Z}; note that subjects were not provided any such figure). Subjects played 10 instances of the game, with each instance differing from the others only in the

[8] Corpus linguistics studies language as expressed in samples of "real world" text, such as the Google Books corpus (which contains the full text of books and magazines that Google has scanned and stored in its digital database).

names of the three options. Each subject was assigned the same partner for all the (10) games. *No feedback* was provided between games.

		Player 2		
		X	**Y**	**Z**
	X	0.10, 0.10	0, 0	0, 0
Player 1	**Y**	0, 0	0.10, 0.10	0, 0
	Z	0, 0	0, 0	0.10, 0.10

Figure 1 - The coordination game. The bottom-left and top-right numbers in each cell represent the monetary payoffs to Player 1 and Player 2, respectively. (For expositional purposes, the set of strategies is denoted by {X, Y, Z}; subjects were not provided any such figure.)

We ran two versions of the study: Version A's options consisted of names of countries, whereas Version B's options consisted of names of food ingredients. The reason we designed two versions is to verify that the presumed prominence of frequently-mentioned labels does not depend on the characteristics of a specific collection. Note that the option sets for Version A were obtained by selecting member states of the United Nations *at random*. The option sets for Version B were obtained by selecting single-word items *at random* from a list of ingredients, which had been scraped from the Epicurious cooking website (https://www.epicurious.com/).

Participants were randomly assigned to the two versions: 48 subjects were assigned to Version A, and 43 subjects to Version B.[9] Average earnings were GBP 0.63 and 0.52 for Versions A and B, respectively, not including subjects' participation fees. Finally note that, in each version, the order of the games was randomized across subjects. By contrast, the order of the three options – in a given game – was determined prior to the experiment at random, and was identical across subjects (i.e., the options were displayed in the same order as in Figure 2 below).

[9] Given the odd number of subjects in Version B, one participant was assigned two partners (but received compensation for playing with either one, at random); the two partners were treated like any other participant.

Version A

	[option X]	[option Y]	[option Z]
1	Kyrgyzstan	Tuvalu	Morocco
			h
2	Turkey	Jamaica	Senegal
	h		
3	Yemen	Benin	Jamaica
			h
4	Spain	Angola	Norway
	h		
5	Bolivia	Kyrgyzstan	Palau
	h		
6	Afghanistan	Germany	Ghana
		h	
7	Nepal	Uzbekistan	Thailand
			h
8	Bahamas	Botswana	Tuvalu
		h	
9	Bahamas	Eritrea	Ghana
			h
10	Ukraine	Jordan	Zambia
		h	

Version B

	[option X]	[option Y]	[option Z]
1	mandarin	soybean	cinnamon
			h
2	raspberry	sauerkraut	scallion
	h		
3	cashew	vanilla	yogurt
		h	
4	meat	cocoa	raspberry
	h		
5	horseradish	rhubarb	tarragon
		h	
6	butter	peppermint	cocoa
	h		
7	pineapple	nutmeg	buckwheat
	h		
8	apple	milk	tuna
		h	
9	buckwheat	tomato	citrus
		h	
10	brandy	pumpkin	tomato
			h

Figure 2 - The option sets for Study 1. The left and right panels refer to Versions A and B, respectively. The blue, maroon and green cells below each strategy-label indicate the option with the relatively highest ("*h*"), middling and lowest NGRAM, respectively.

Results. The average participant in Study 1 chose the strategy associated with the highest, middling and lowest NGRAM respectively 45.27%, 39.34% and 15.39% of the time (in particular, the average participant chose the option with the *highest* NGRAM 45.83% of the time in Version A, and 44.65% of the time in Version B).[10] This provides some preliminary evidence

[10] Please refer to Appendix A for a bar graph of the frequency distributions of individual-level choices for each of the games.

about the impact of frequently-mentioned labels on strategic play. We stress that such a distribution significantly differs from play in the fully mixed equilibrium assigning equal probability to each strategy (N = 91 obs., T^2 = 127.67, p = 0.000 under a Hotelling's T-squared generalized means test conducted on the *sample of per-subject mean choices*).[11] In this regard, we note that – since any three options constitute symmetric strategies – here Harsanyi and Selten (1988) would argue that the only "symmetry-invariant" equilibrium must assign them each the same probability (in order for a renaming of strategies not to affect the solution of the game). Their prediction is clearly falsified by the data. Furthermore, we stress that the fully mixed equilibrium would imply an expected coordination rate of 0.33, whereas the expected frequency of coordination resulting from our sample of (per-subject) mean observations is 0.57 for Version A, and 0.49 for Version B.[12,13]

One more comment is in order. We designed Study 1 as a preliminary test for predicting behavior in coordination games: the method of analysis employed so far has utilized mean observations, thereby discarding a substantial amount of information. That said, in what follows we introduce a set of between-subjects designs to explicitly test for the strategic use of word frequency: there we shall perform a more sophisticated analysis in such a way to account for the characteristics of *each game* (i.e., option triplet) to which *each subject* was exposed.

---

[11] The sample of (per-subject) mean observations is obtained as follows. First, for each choice of subject i, assign a value of 1 or 0 to indicate if the option with the *highest* NGRAM was chosen or not; then, take the average value across all the choices of i. Similarly, assign a value of 1 or 0 to indicate if the option with the *middling* NGRAM was chosen or not, and take the average value across all the choices of i. Lastly, assign a value of 1 or 0 to indicate if the option with the *lowest* NGRAM was chosen or not, and take the average value across all the choices of i.

[12] In keeping with previous studies, here we focus on expected (as opposed to actual) coordination rates. Actual coordination rates depend on individual choices and on a purely random element, that is, the random assignment of partners. It follows that – given a relatively small sample size – such a random element is likely to bias coordination rates. «Thus, the *actual* frequency of coordination has no special significance; it is more appropriate to consider the *expected* frequency of coordination» (Mehta et al., 1994, p. 663, italics in original).

[13] This rate measures the probability that two individuals – selected at random from the pool of participants in Study 1 – select the same strategy in a randomly chosen question. Here, we calculate this rate using Monte Carlo methods. The pseudo code is as follows: (A) Pick two participants at random from the sample; (B) Pick one of the ten games at random; (C) If both participants chose the same strategy in the game consider it an instance of successful coordination, otherwise consider it unsuccessful; (D) Repeat steps A-C 100,000 times to calculate the average frequency of successful coordination.

IV - Study 2

Demographics. The subject pool for Study 2 consisted of 160 US resident individuals. The average participant was 30 years old, and 58% of the subjects were male. As with Study 1, participants took less than 10 minutes to review the instructions and complete all the tasks. Subjects received a 0.5 GBP participation fee, in addition to the payoffs earned in each game (if any), as specified below.

Design. The goal of the study was to test for the strategic use of word frequency by systematically varying players' incentives. To that end, we designed a few variants of the simple coordination game of Study 1, in such a way to incentivize or disincentivize coordination for either player. To test for replicability, Study 2 featured the exact same option triplets as in Version A of Study 1—see the left panel of Figure 2 above. Study 2 involved the following four roles/conditions.

a. ***Coordinate***: This is a replication of Version A of Study 1 that, among the other purposes, served to verify the robustness of our earlier results. Participants in this condition ("Coordinators") were paired with other participants in the same condition, and were informed so.
b. ***Pick***: Participants in this condition ("Pickers") were presented with the same labels as in the Coordinate condition, except that they were asked to merely pick one of the three given options. That is, participants were not assigned a partner, nor did they receive any additional payoffs on the basis of their choices; hence, they had no strategic incentives to select one label over another.
c. ***Seek***: This condition featured the same strategy-labels as in the Coordinate condition, except that the incentive structure reflected the role of "Seeker" in the Hide-and-Seek game: Figure 3 below represents the payoff structure of this game. As can be seen there, a Seeker receives GBP 0.10 if both members of a pair choose the same strategy, and nothing otherwise. Participants in this condition were paired with participants in the Hide condition below, and were informed so.
d. ***Hide***: Again, this featured the same strategy-labels as in the Coordinate condition, except that the incentive structure reflected the role of "Hider" in the Hide-and-Seek game. As can be seen in Figure 3, a Hider receives GBP 0.10 if members of a pair choose different

strategies, and nothing otherwise. Participants in this condition were paired with participants in the Seek condition above, and were informed so.

Each subject completed 10 tasks, in the same role/condition, with each task differing from the others only in the names of the three options (see the left panel of Figure 2 above). In the Coordinate, Seek, and Hide conditions each subject was assigned the same partner for all the (10) games, and was informed so. *No feedback* was provided between games.

		Seeker		
		X	**Y**	**Z**
	X	0.10 0	0 0.10	0 0.10
Hider	**Y**	0 0.10	0.10 0	0 0.10
	Z	0 0.10	0 0.10	0.10 0

Figure 3 - The Hide-and-Seek game structure in bimatrix form. The bottom-left and top-right numbers in each cell represent the payoffs to the *Hider* and *Seeker*, respectively.

Since all our conditions involved the exact same options, if the effect of word frequency in Study 1 were merely due to an automatic (or naïve) response, then we should observe similar choice distributions across conditions. If however participants used word frequency in a strategic manner, then we should find that frequently-mentioned labels are more (less) likely to be selected when there are incentives to match (mismatch) others, with the magnitude of the change varying with subjects' strategic sophistication.

The intuition is rationalized by specifying a level-*k* reasoning model (Nagel, 1995; Stahl and Wilson, 1995; Costa-Gomes, Crawford, and Broseta, 2001; Camerer, Ho, and Chong, 2004; Costa-Gomes and Crawford, 2006). Level-*k* theories assume a hierarchy of player types defined by the level of sophistication with which each player reasons, with players' beliefs being based on simple non-equilibrium models of others (Crawford, Costa-Gomes, and Iriberri, 2013). More precisely, level-*k* types anchor their beliefs in a non-strategic *L0* type and adjust them via iterated best responses, so that *L1* best responds to *L0*, *L2* best responds to *L1*, and so on.[14] In what follows we operationalize this approach by specifying a set of assumptions that will generate our predictions.

(i) A non-strategic *L0* type in the *Coordinate*, *Seek*, or *Hide* conditions behaves like a participant in the *Pick* condition.[15]

(ii) Some *L1* players believe that the distribution q of *L0* choices is uniformly random, whereas other *L1* players believe that the distribution q of *L0* choices is *not* uniformly random.

(iii) There are no players at *L3* or higher.

A few comments are in order.

With regard to assumption (i), we note that all level-*k* variants universally posit that *L0* types do not engage in strategic reasoning, but simply randomize between options according to some probability distribution q (Crawford et al., 2013). But what determines such a distribution? In the models of Stahl and Wilson (1995) and Camerer et al. (2004), that distribution is assumed to be uniformly random. Other applications of level-*k* reasoning (such as Crawford and Iriberri, 2007) instead assume that the distribution is not uniform in that *L0* types are more likely to choose "salient" (i.e., prominent) labels, without however defining what salience is. Now, in our case *if* one were to assume that *L0* choices are not uniformly random, then a plausible deviation from a uniform distribution may be that the probability of selecting the most frequently-mentioned label (i.e., the option with the highest NGRAM) at *L0* is higher than the probability of selecting any other label.[16]

[14] We note that different applications of level-*k* reasoning make different assumptions as to whether there are actually any players at *L0*. Applications also differ in their assumptions as to the players' depth of reasoning about other types; that is, in some cases it is assumed that players at each level above *L0* best respond to a probability mix of the decisions of all levels below their own (as opposed to best responding to the one level immediately below). For discussion see Crawford, Costa-Gomes, and Iriberri (2013). We finally note that, in our context, such differences do not result in different behavioral predictions.

[15] This would apply even to models assuming that the relative frequency of *L0* types is zero, in the sense that – in that case – *L1* types would best respond to their beliefs about potential *L0* behavior, with such beliefs reflecting Pickers' behavior.

[16] As noted above, research in cognitive psychology has shown that the frequency of occurrence of a word in everyday language is positively correlated with the tendency to select that word as a response in *non-strategic* tasks (Dougherty et al., 2008; Goldstein and Gigerenzer, 2002; Schooler and Hertwig, 2005).

Here, we partly sidestep the issue of specifying the probability distribution q by empirically defining it on the basis of Pickers' behavior. Given this, in principle one could formalize a level-*k* model by assuming that players at *L1* or above best respond to a commonly known distribution corresponding to q. However, in the context of one-shot coordination games that assumption appears very implausible (Bardsley, Mehta, Starmer, and Sugden, 2010, p. 45). It follows that the rational course of action for players at *L1* or above ultimately depends on their guesses about *L0* behavior. Since we do not observe such guesses, we formulated assumption (ii) above: so we posit that some *L1* players believe that the distribution of *L0* choices is uniformly random (as per Stahl and Wilson, 1995, and Camerer et al., 2004), and we posit that some other *L1* players believe that *L0* choices are *not* uniformly random (as per Crawford and Iriberri, 2007). More specifically, we assume that the latter believe that one deviation from a uniform distribution is that "the probability of *L0* players selecting the most frequently-mentioned label is higher than any other label". (This line of analysis is extended to higher-level types, by assuming that some *L2* types believe that *L1* players best respond to a uniform distribution, whereas other *L2* types believe that *L1* players best respond to a non-uniform distribution, and so on.)

We finally note that (iii) is a simplifying assumption, based on previous empirical evidence on subjects' strategic sophistication. For example, Arad and Rubinstein (2012) observed that level-*k* experiments have shown that «the most frequent types are usually *L1* and *L2*, whereas higher-level types are rare» (p. 3561). For a similar point, see also Penczynski's (2016) analysis of Hide-and-Seek games.

What do our assumptions imply in terms of behavioral predictions? In short, in the *Coordinate* condition, some *L1* types will randomize uniformly across labels while some other *L1* types will select (with probability one) the label that is believed to be the modal choice at *L0*, namely, the option with the highest NGRAM. The same prediction applies to higher-level types. Hence, depending on the relative frequency of those two *L1* types, it follows that:

***H1**: Coordinators select the most frequently-mentioned label as many times as – or more than – Pickers.*

Next, without making any further assumptions about the specific distribution of types, for the purposes of this study we limit ourselves to verifying whether the effect of word frequency is strategic. If it were fully non-strategic, then we should observe the same choice distributions across all conditions. Instead if participants used word frequency in a boundedly rational manner, then based on the assumptions above we should observe that:

***H2**: Coordinators are (weakly) more likely to select the most frequently-mentioned label than Seekers.*

***H3**: Hiders are (weakly) less likely to select the most frequently-mentioned label than Pickers or Seekers.*

Formally, denote by q_H the probability of selecting the most frequently-mentioned label at *L0*. Assume that a fraction x_1 of *L1* types believe that the behavior of *L0* players is described by a random uniform distribution, and therefore believe that q_H is $\frac{1}{3}$. On the other hand, assume that a fraction y_1 of *L1* types believe that q_H is higher than the probability of selecting any other label. Finally assume that a fraction x_2 of *L2* types best respond to x_1, whereas a fraction y_2 of *L2* types best respond to y_1. Given this, the assumptions above entail that the Coordinators' average probability p_{COORD} to select the most frequently-mentioned label across levels is defined by $p_{COORD} = l_0 \cdot q_H + l_1 \cdot \left(\frac{1}{3} \cdot x_1 + 1 \cdot y_1\right) + l_2 \cdot \left(\frac{1}{3} \cdot x_2 + 1 \cdot y_2\right)$, where l_k denotes the total share of *Lk* players in the population, while the first and second expressions in parentheses denote the behavior of *L1* and *L2* Coordinators, respectively. Further, the Seekers' average probability p_{SEEK} to select the most frequently-mentioned label across levels is given by $p_{SEEK} = l_0 \cdot q_H + l_1 \cdot \left(\frac{1}{3} \cdot x_1 + 1 \cdot y_1\right) + l_2 \cdot \left(\frac{1}{3} \cdot x_2 + 0 \cdot y_2\right)$, where the first and second expressions in parentheses denote the behavior of *L1* and *L2* Seekers, respectively. Lastly, the Hiders' average probability p_{HIDE} to select the most frequently-mentioned label across levels is given by $p_{HIDE} = l_0 \cdot q_H + l_1 \cdot \left(\frac{1}{3} \cdot x_1 + 0 \cdot y_1\right) + l_2 \cdot \left(\frac{1}{3} \cdot x_2 + 0 \cdot y_2\right)$, where the first and second expressions in parentheses denote the behavior of *L1* and *L2* Hiders, respectively. It follows that $p_{COORD} \geq p_{SEEK} \geq p_{HIDE}$ and $p_{COORD} \geq p_{PICK} \geq p_{HIDE}$, with $p_{PICK} = q_H$.[17]

[17] By definition *L0* behavior is the same across roles: this means that the probability that *L0* Hiders and *L0* Seekers select the most frequently-mentioned label is given by q_H (regardless of the specifics of the distribution q). Also note that level-*k* reasoning entails that Seekers select the label that is believed to be the modal choice of opponents at the level below, whereas Hiders want to avoid such a label. Here, this means that the most frequently-mentioned label will be selected by (y_1) *L1* Hiders and (y_1) *L1* Seekers with probability zero and one, respectively; on the other hand, (x_1) *L1* Hiders and (x_1) *L1* Seekers will randomize uniformly. Iterated best responses are similarly implemented by higher-level players: (y_2) *L2* Hiders and (y_2) *L2* Seekers will select the most frequently-mentioned label with probability zero; on the other hand, (x_2) *L2* Hiders and (x_2) *L2* Seekers will randomize uniformly.

Results. Table 1 presents mean choices in each of the four conditions, given a classification of the strategy-labels based on their relative n-gram frequency. (Please refer to Appendix A for a bar graph of the frequency distributions of individual-level choices for each of the games and conditions.) By giving a glance at Table 1 below, the reader will easily notice that the mean distribution of choices varied with each condition. In particular, the most frequently-mentioned option (i.e., the strategy-label with the highest NGRAM) was selected less and less often when moving from *Coordinate* (or *Pick*) to *Seek* to *Hide*: a Kruskal-Wallis test on the choice of such frequently-mentioned labels confirms significant differences across conditions (note that to satisfy the assumption of independence of observations, all our non-parametric tests are conducted on the sample of per-subject mean choices; N = 160 obs., χ^2_3 = 10.477, p = 0.014, two-tailed). On the other hand, it is noteworthy that the least frequently-mentioned option (i.e., the strategy-label with the lowest NGRAM) was selected more and more often when moving from *Coordinate* to *Seek* to *Hide*: a Kruskal-Wallis test on the choice of such "infrequently-mentioned" labels confirms significant differences across conditions (N = 160 obs., χ^2_3 = 34.353, p = 0.000, two-tailed).

Following on from the above discussion, we proceed to test some directional hypotheses (via simple pairwise comparisons) by using the sample of per-subject mean choices. We begin by contrasting choice behavior in the Pick and Coordinate conditions: a one-tailed test allows us to examine the *alternative hypothesis* that the most frequently-mentioned option (i.e., the strategy-label with the highest NGRAM) is selected strictly less often in Coordinate than in Pick.[18] In short, a one-tailed Wilcoxon-Mann-Whitney shows no evidence of a significant decrease in the choice of the most frequently-mentioned label when moving from *Pick* to *Coordinate* (N = 81 obs., Z = -1.008, p = 0.1567). This provides some very preliminary evidence in support of H1.

[18] Recall that H1 says that Coordinators select the most frequently-mentioned label as many times as – or more than – Pickers. Because of the weak inequality, here we test against the alternative hypothesis that the most frequently-mentioned label is selected strictly less often in *Coordinate* than in *Pick*.

Choice by word frequency	***Coordinate***	***Pick***	***Seek***	***Hide***
Strategy-label with *highest* NGRAM metric is chosen [f_H], %	46.19 (.1464)	48.46 (.1646)	43.64 (.1556)	36.86 (.1548)
Strategy-label with *middling* NGRAM metric is chosen [f_M], %	48.10 (.1596)	38.21 (.1211)	34.09 (.1661)	32.85 (.1202)
Strategy-label with *lowest* NGRAM metric is chosen [f_L], %	5.71 (.0914)	13.33 (.1675)	22.27 (.2044)	30.29 (.2121)
Total, %	100	100	100	100
Total # triplets (1,600) i.e., no. of triplets * subjects	420	390	440	350
Subjects (160)	42	39	44	35

Table 1 - (Per-subject) mean choice, given a classification of the strategy-labels based on the NGRAM metric; in brackets is the standard deviation.[19]

Whereas a similar analysis shows no significant difference between Coordinators' and Seekers' behavior, we find that the most frequently-mentioned label was selected significantly less in *Hide* than in *Pick*, providing some preliminary support for H3 ($N = 74$ obs., $Z = 3.101$, $p = 0.000$ under a one-tailed Wilcoxon-Mann-Whitney test). Similarly, the most frequently-mentioned label was selected significantly less in *Hide* than in *Seek* ($N = 79$ obs., $Z = 1.978$, $p = 0.023$, one-tailed Wilcoxon-Mann-Whitney test), which again supports H3. In summary, these non-parametric tests show that the most frequently-mentioned label was selected less often when moving from *Pick* to *Hide* and from *Seek* to *Hide*. We however note that – since the tests above are conducted on the sample of per-subject mean choices – these tests do not account for differences in individual responses across games, or differences in the relative magnitude of the n-gram frequency across games. Moreover, the tests above do not control for the labels' position on the screen (i.e., top, middle, and bottom of the list).

[19] Given that the number of subjects in *Hide* was less than in *Seek*, for the mere purpose of calculating Seekers' payoffs, nine Hiders were matched with two Seekers (but received compensation for playing with either one, at random).

For these reasons, we corroborated our analysis by running alternative-specific conditional logistic regressions (*asclogit*; McFadden, 1973). The results of such regressions shall provide the ultimate test of our hypotheses, by accounting for the characteristics of each game (i.e., option triplet) to which each subject was exposed. Below we report the main findings while we refer the reader to Appendix A for the full regression tables.

Overall, the regressions confirm a positive and significant effect of word frequency on choice, across treatments: in plain words, the higher the n-gram frequency of a label, the more likely it is for the associated strategy to be selected, regardless of any other characteristics (e.g., experimental condition or label's position on the screen). In particular, when contrasting choice behavior in *Pick* against *Coordinate*, the regression reveals no significant difference in the relative impact of the n-gram frequency between these two conditions: that is, we find no difference in the probability of choosing the most frequently-mentioned label between Pickers and Coordinators, confirming our previous result (see regressor *WFC* in model [1] of Table 1A, in Appendix A). Next, when contrasting choice behavior in *Coordinate* against *Seek*, we find a significant difference in the relative impact of the n-gram frequency on choice: that is, a label with a higher n-gram frequency is more likely to drive the choices of Coordinators than Seekers (coef. = -.119, z = -2.44, p = 0.015, two-tailed asclogit conducted on the sample of individual observations, with standard errors adjusted for clustering on 86 subjects; see regressor *WFC* in model [2] of Table 1A). The result clearly supports H2.

Further, contrasting choice behavior in *Pick* against *Hide*, we find a significant difference in the relative impact of the n-gram frequency on choice: a label with a higher n-gram frequency is more likely to drive the choices of Pickers than Hiders (coef. = -.201, z = -3.77, p = 0.000, two-tailed asclogit conducted on the sample of individual observations, with standard errors clustered on 74 subjects; see regressor *WFC* in model [3] of Table 2A). Moreover, contrasting choice behavior in *Seek* against *Hide*, we find again a significant difference in the relative impact of the n-gram frequency on choice: that is, a label with a higher n-gram frequency is more likely to drive the choices of Seekers than Hiders (coef. = -.2145, z = -2.14, p = 0.033, two-tailed asclogit conducted on the sample of individual observations, with standard errors clustered on 79 subjects; see regressor *WFC* in model [4] of Table 2A). The results clearly support H3.

To conclude, the data provide strong support for our hypotheses (please refer to Appendix A for an extended commentary on the regression analysis). In particular, despite the fact that all conditions involved exactly the same option triplets, more frequently-mentioned labels were selected less often when moving from *Coordinate* to *Seek*, from *Pick* to *Hide*, and from *Seek* to *Hide*. These data patterns suggest that individuals do select strategies that fulfill our characterization of prominence, and they do so in a (boundedly) rational manner consistent with our level-*k* model specification. In the remainder of the article we shall delve further into the strategic use of labels in coordination games.

V - Study 3

Demographics. The subject pool for Study 2 consisted of 80 individuals, of which half were US residents and half were UK residents. All participants were recruited from the Prolific Academic website, simultaneously. In the US-residents sample, the average participant was 29 years old and 75% of subjects were male. In the UK-residents sample, the average participant was 31 years old and 63% of subjects were male. As with our previous studies, participants took less than 10 minutes to review the instructions and complete all the tasks. Subjects received a 0.5 GBP participation fee, in addition to the payoffs earned in each game, as specified below.

Design. In what follows we present a set of treatments designed to delve into the strategic use of labels in one-shot coordination games. To that end, Study 3 involved the same pure coordination structure as in Study 1 (see payoff table in Figure 1 above), except that here we used a different set of strategy-labels. So unlike our previous studies (where we used randomly-generated labels), for this study we purposely selected triplets of labels in such a way that the option with the highest word frequency varied between the American- and British-English vocabularies (as measured by the relevant NGRAM in the American- and British-English Google Books corpora).

Specifically, subjects played 10 instances of coordination games, with each instance differing from the others only in the names of the three options. Each subject was assigned the same partner for all the (10) games, and was informed so. *No feedback* was provided between games. The order of the games was randomized across subjects. By contrast, the order of the three options – in a given game – was determined prior to the experiment at random, and was identical across subjects (i.e., the options were displayed in the same order as in Figure 4).

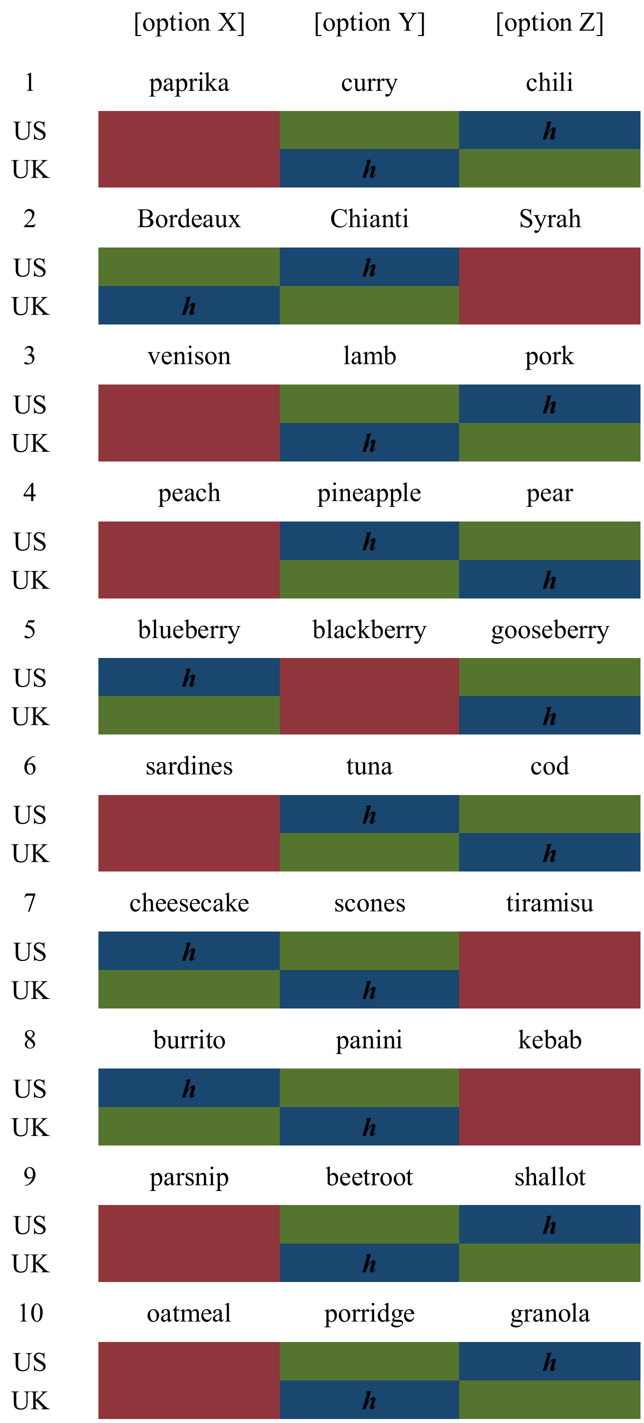


Figure 4 - The option sets for Study 3. The blue, maroon and green cells in the first (*second*) row below each strategy-label respectively indicate the option with the relatively highest ("*h*"), middling and lowest American-English (*British-English*) NGRAM.

As noted above, in this study we varied the cultural makeup of the subject pool (by recruiting a sample of US residents and a sample of UK residents). Furthermore, we manipulated participants' perception of their counterpart's cultural/linguistic background: more specifically, we manipulated the information provided to each subject in regards to their partner's country of residence. In short, each participant – whether in the US or in the UK sample – was assigned to one of the following information conditions.

a. "**NO-k**": Participants in this condition received no information about their partners' residence (hence, this condition is identical to Study 1, except that we used different strategy-labels). The instructions in this condition read: *"Please choose one option. Each of you and your partner receive £0.10 if you both choose the same option, £0 otherwise."*
b. "**k-UK**": Participants in this condition were told that their partner resided in the UK. Specifically, they were shown the following message: *"Please choose one option. Each of you and your partner receive £0.10 if you both choose the same option, £0 otherwise. Your partner is a Prolific worker who resides in the UK. Your partner may or may not know where you reside."*
c. "**k-US**": Participants in this condition were told that their partner resided in the US. Specifically, they were shown the following message: *"Please choose one option. Each of you and your partner receive £0.10 if you both choose the same option, £0 otherwise. Your partner is a Prolific worker who resides in the US. Your partner may or may not know where you reside."*

In summary, in both the US and UK samples, each subject was assigned to one of the above three conditions. This means that our study involved a total of six sub-conditions: 1) US participants without additional knowledge about their partner; 2) UK participants without additional knowledge about their partner; 3) US participants knowing that they were matched with a US participant; 4) US participants knowing that they were matched with a UK participant; 5) UK participants knowing that they were matched with a US participant; 6) UK participants knowing that they were matched with a UK participant.

In terms of predictions we stress that – since all conditions involved the same options – if the effect of word frequency in previous coordination games were merely due to an automatic (or naïve) response, then here we should observe similar choice distributions across conditions. If however participants used word frequency in a strategic manner, then we should find that frequently-mentioned labels are more likely selected when subjects think that their assigned partner may be culturally alike (and hence view the problem in a similar way). In fact, if there is reason to believe that an option easily comes to mind to people who are culturally alike, then it might be optimal to select that option in coordination games. We note that this intuition is consistent with "team reasoning", whereby individuals follow a rule of selection which, if followed by all other (similar) players, would be optimal for each of them (Sugden, 1995). Moreover, we note that this intuition would be consistent also with a level-*k* model specification allowing for multiple *L0* types (such that each *L0* type selects the label that is most frequently mentioned in their own vocabulary, and players at *L1* or above best respond to their guesses about *L0* behavior).

***H4**: Choice behavior differs between US and UK residents, and in each case it is positively related to the word frequency of an option (as measured by the n-gram frequency in the relevant vocabulary).*

***H5**: Subjects who are informed that their counterpart resides in the same country are more likely to select the most frequently-mentioned label (in their vocabulary) than subjects who are unaware of the counterpart's residence; in turn, the latter are more likely to select the most frequently-mentioned label (in their vocabulary) than subjects who are informed that their counterpart resides in a different country.*

Results. Table 2 presents mean choices in each of the six sub-conditions, given a classification of the strategy-labels based on the relevant n-gram frequency. (For the distribution of individual-level choices in each of the 10 games, please refer to Figure 3A: *US sample* and Figure 4A: *UK sample*, in Appendix A.) By giving a glance at Table 2 below, the reader will notice that the distribution of choices clearly varied with each sub-condition.

Next, for ease of exposition Table 3 below pools our six sub-conditions into three samples. The first sample ("k-SAME") consists of participants who were informed that their assigned partner resided in the same country (i.e., US participants in the *k-US* condition, and UK participants in the *k-UK* condition). The second sample consists of participants who were not informed of their partner's residence, and corresponds to all participants in the *NO-k* condition. The third sample ("k-OTHER") consists of participants who were informed that their assigned partner resided in a different country (i.e., US participants in the *k-UK* condition, and UK participants in the *k-US* condition). In what follows we report the results of non-parametric tests conducted on such a pooled (US & UK) sample, whereas later on we discuss each of the US and UK samples in turn.

	US sample			UK sample		
Choice by word frequency	***NO-k***	***k-UK***	***k-US***	***NO-k***	***k-UK***	***k-US***
Strategy-label with *highest* NGRAM metric is chosen [f_H], %	47.89 (.1812)	35.00 (.2173)	57.27 (.1618)	35.00 (.1617)	43.33 (.1775)	26.00 (.1577)
Strategy-label with *middling* NGRAM metric is chosen [f_M], %	31.05 (.1663)	32.00 (.1475)	27.27 (.1902)	23.33 (.1533)	30.83 (.1378)	25.00 (.1715)
Strategy-label with *lowest* NGRAM metric is chosen [f_L], %	21.06 (.1559)	33.00 (.1828)	15.46 (.0934)	41.67 (.1617)	25.84 (.1443)	49.00 (.1663)
Total, %	100	100	100	100	100	100
Total # triplets i.e., no. of triplets * subjects	190	100	110	180	120	100
Subjects	19	10	11	18	12	10

Table 2 - (Per-subject) mean choice, given a classification of the strategy-labels based on the relevant NGRAM metric (American-English for *US sample* and British-English for *UK sample*); in brackets is the standard deviation.

As can be seen in Table 3 below, the average participant in k-SAME chose the strategy associated with the highest, middling and lowest NGRAM (assessed with reference to the subject's own vocabulary) respectively 50.00%, 29.13% and 20.87% of the time. By contrast, the average participant in NO-k chose those strategies 41.62%, 27.29% and 31.09% of the time, respectively.

Choice by word frequency	US & UK samples		
	k-SAME	***NO-k***	***k-OTHER***
Strategy-label with *highest* NGRAM metric is chosen [f_H], %	50.00 (.1809)	41.62 (.1818)	30.50 (.1904)
Strategy-label with *middling* NGRAM metric is chosen [f_H], %	29.13 (.1621)	27.29 (.1627)	28.50 (.1598)
Strategy-label with *lowest* NGRAM metric is chosen [f_H], %	20.87 (.1311)	31.09 (.1882)	41.00 (.1889)
Total, %	100	100	100
Total # triplets i.e., no. of triplets * subjects	230	370	200
Subjects	23	37	20

Table 3 - (Per-subject) mean choice, given a classification of the strategy-labels based on the relevant NGRAM metric (American-English for *US sample* and British-English for *UK sample*); in brackets is the standard deviation. NOTE: *k-SAME* contains observations from US subjects in k-US condition, as well as UK subjects in k-UK condition; *k-OTHER* contains observations from US subjects in k-UK condition, as well as UK subjects in k-US condition.

It is clear that the distribution of choices in each of the *k-SAME* and *NO-k* samples differed from the fully mixed equilibrium assigning equal probability to all strategies, as is confirmed by Hotelling's T-squared generalized means tests conducted on the samples of per-subject mean choices (for k-SAME: N = 23 obs., T^2 = 26.91, p = 0.000; for NO-k: N = 37 obs., T^2 = 9.21, p = 0.018). On the other hand, there was no such difference in the distribution of choices in *k-OTHER* (N = 20 obs., T^2 = 3.73, p = 0.199, Hotelling's T-squared generalized means tests). The *k-OTHER* pattern shows that only subjects who did not have a compelling reason to rely on word frequency behaved as postulated by Harsanyi and Selten (1988); that is, they assigned each (symmetric) strategy the same probability.

Overall, the data suggest that participants used word frequency as a cue in k-SAME and NO-k, but not in k-OTHER. In fact, the option with the highest word frequency (i.e., the strategy-label with the highest NGRAM in the subject's own vocabulary) was selected less and less often when moving from *k-SAME* to *NO-k* to *k-OTHER*: a Kruskal-Wallis test on the choice of the most frequently-mentioned label confirms significant differences across these three samples (N = 80 obs., χ_3^2 = 10.597, p = 0.005, two-tailed). This provides some preliminary evidence in support of H5.

	US sample	UK sample
Choice by location		
Top of list is chosen, %	53.75 (.2723)	40.50 (.2363)
Center of list is chosen, %	26.00 (.1864)	39.75 (.2118)
Bottom of list is chosen, %	20.25 (.1544)	19.75 (.1367)
Total, %	100	100
Total # triplets i.e., no. of triplets * subjects	400	400
Subjects	40	40

Table 4 - (Per-subject) mean choice, given a classification of the strategy-labels based on the labels' position on the screen.

Table 4 provides a different view of the data by categorizing mean choices with respect to the labels' *position on the screen* (i.e., top, middle, and bottom of the list). Note that the order of the three labels – in a given game – was identical across subjects and countries (i.e., the options were displayed in the same order as in Figure 4 above). Hence, if there were no causal relationship between word frequency and choice behavior, then choice distributions should be similar; that is, US participants should be as likely as UK participants to choose the option displayed at the top of the list on the screen (the same argument applies to the probability of choosing middle and bottom options, across countries). However, it is clear that the *US and UK choice distributions differed from each other* (N = 80 obs., T^2 = 9.62, p = 0.011 under a two-group Hotelling's T-squared generalized means test conducted on the samples of per-subject mean choices). This provides some preliminary evidence in support of H4.

The non-parametric tests we have presented so far provide a cursory overview of behavioral patterns in our (per-subject mean) dataset. Such tests however disregard any differences in individual responses across games. For this reason, we corroborated our findings by conducting a regression analysis of the full sample of individual observations (while adjusting standard errors for clustering on the subjects). We begin by formally investigating the relationship between one's use of word frequency and one's knowledge of the counterpart's country of residence. To that end, we coded a *"knowledge"* ordinal variable as follows: *knowledge* takes on value 1 if a subject was in the k-OTHER sample (i.e., one is informed that the assigned partner resides in a different country); *knowledge* takes on value 2 if a subject was in NO-k (i.e., one receives no information about the partner's country of residence); *knowledge* takes on value 3 if a subject was in the k-SAME sample (i.e., one is informed that the partner resides in the same country). We then coded a binary *"successful prediction"* variable, which takes on value 1 if a subject selected the most frequently-mentioned label (as measured by the NGRAM in the relevant vocabulary), and takes on value 0 otherwise. Given that, we simply regressed the prediction variable on the knowledge level. In short, in the *US sample* we found a positive and significant effect of knowledge on prediction (coef. = .4529, $z = 2.65$, $p = 0.008$, two-tailed logit with standard errors clustered on 40 subjects). Unsurprisingly, we found a similar effect in the *UK sample* (coef. = .3851, $z = 2.44$, $p = 0.015$, two-tailed logit with standard errors clustered on 40 subjects). This confirms that our subjects used word frequency in a strategic manner: the more they had reason to believe that their assigned partner was alike, the more likely they were to select the most frequently-mentioned label in their own vocabulary. The results provide striking support for H5.

We move on to present some robustness checks. In order to control for differences in the relative magnitude of the n-gram frequency across games and vocabularies – as well as differences in the labels' position on the screen – we ran some alternative-specific conditional logistic regressions (*asclogit*; McFadden, 1973). These regressions account for the characteristics of each option triplet to which each subject was exposed. Below we report the main findings while we refer the reader to Appendix A for the full regression tables.

In brief, the regressions provide further evidence in support of the hypothesis that if one is informed that the assigned partner resides in a different country, one will less likely choose the strategy associated with the most frequently-mentioned label (regardless of the label's position on the screen). In particular – in the case of the *US sample* – when contrasting choice behavior in *k-US* against *k-UK*, we find a significant difference in the impact of the relevant NGRAM on choice. That is, a label with a higher American-English n-gram frequency is more likely to drive the choices of US participants who know that their partner resides in the US rather than in the UK (coef. = -4.316, z = -2.93, p = 0.003 under a two-tailed asclogit, with standard errors clustered on 21 subjects; see regressor *WFC* in model [5] of Table 3A).

Furthermore – in the case of the *UK sample* – when contrasting choice behavior in *k-UK* against *k-US*, like before we find a significant difference in the impact of the relevant NGRAM on choice. That is, a label with a higher British-English n-gram frequency is more likely to drive the choices of UK participants who know that their partner resides in the UK rather than in the US (coef. = -4.218, z = -2.94, p = 0.003 under a two-tailed asclogit, with standard errors clustered on 22 subjects; see regressor *WFC* in model [6] of Table 3A).

We conclude by discussing expected coordination rates. For this purpose, we compared coordination rates that would be obtained if different subsamples were paired with each other, using Monte Carlo methods.[20] Whereas chance would imply an expected coordination rate of 0.33, our subjects' choice behavior implies expected coordination rates of 0.55 in the *US* k-US condition (i.e., US participants knowingly paired with other US participants) and 0.43 in the *UK* k-UK condition (i.e., UK participants knowingly paired with other UK participants). Notably, these rates drop to 0.47 in the case of *US participants unknowingly paired with other US participants*, and to 0.39 in the case of *UK participants unknowingly paired with other UK participants*. Thus, if a subject knew that her partner was from the same country, the likelihood of successful coordination increased. We finally examined coordination rates when US and UK participants were knowingly paired with each other (i.e., the *US* k-UK sample paired with the *UK* k-US sample) and unknowingly paired with each other (i.e., the *US* NO-k sample paired with the *UK* NO-k sample), and found that the coordination rates were 0.37 and 0.38 respectively.

[20] For the case in which both participants belonged to the same sample (e.g., both were US residents in the k-US condition) the code for computing expected coordination rates was identical to that used for Study 1 (see footnotes 12, 13 above). In the case in which the participants belonged to different samples (e.g., US residents in the k-UK condition and UK residents in the k-US condition), step A was modified so that two participants were picked at random from each sample.

In summary, our results show that successful coordination is more likely when individuals are knowingly paired with partners from their own country, as opposed to when they are knowingly or unknowingly paired with partners from a different country.

VI - Conclusion

We have presented a set of studies designed to test for the strategic use of word frequency in experimental games. In the first study we found that the relative frequency with which labels are mentioned in natural language (as measured by their frequency of occurrence in the Google Books corpus) is a good predictor of the likelihood that participants select the corresponding strategies in coordination games. Our second study aimed to verify whether subjects use frequently-mentioned labels strategically rather than naïvely. To do so, we contrasted the behavior of subjects participating in a matching game ("Coordinators") with the behavior of subjects in three alternative roles/conditions, namely, "Pickers", "Hiders", and "Seekers". Our data show that Pickers were about as likely to select the most frequently-mentioned label as were Coordinators; by contrast, Hiders were less likely to select the most frequently-mentioned label than Seekers and, in turn, Seekers were less likely than Coordinators. This pattern suggests a (boundedly) rational use of frequently-mentioned labels. Lastly, we designed our third study to delve further into the strategic use of word frequency in pure coordination games: to that end, we recruited subjects residing in either the US or the UK, and then varied their knowledge of the counterpart's country of residence. Consistent with our predictions, the data show that choice behavior differed between US and UK participants, and in each case it was positively related to the word frequency of an option, as measured by the n-gram frequency in the relevant vocabulary. Moreover, subjects were less likely to rely on word frequency as a means to guiding their behavior when they knew that their counterpart resided in a different country (and thus had a different cultural/linguistic background).

Our results confirm that – ceteris paribus – players find particularly prominent (i.e., salient) strategies denoted by labels frequently used in everyday language. Here we focused on symmetric games, in order to rule out frame or payoff asymmetries as an explanation for the focality of a solution (besides, our focus on *pure* coordination games clearly rules out any confounds due to social preferences). Yet our approach may have wider application. We note that our characterization of prominence has the merit of allowing for clear-cut, quantifiable tests:

word frequency is easily measured in large-scale text corpora, and can be used as a proxy for prominence in games with *any* natural-language label. In this regard, we stress that whereas previous research investigated the role of "positively-valenced" labels in inducing norms (in social dilemmas),[21] no prior paper has systematically quantified the impact of word frequency on the attractiveness of a strategy-label.

Our approach has been motivated by prior evidence on the use of word frequency as a cue in non-strategic tasks (Dougherty et al., 2008; Goldstein and Gigerenzer, 2002; Schooler and Hertwig, 2005). Indeed, research from cognitive psychology has established that the frequency of exposure to words is closely related to "word fluency", that is, the ease with which an individual is able to recognize, retrieve, and process a word. In this paper, we have shown that word frequency has an important role in strategic reasoning as well.

Finally, our studies provide evidence in support of the idea that the prominence of a solution often depends on players' culture. Most notably, our results suggest that individuals are consciously *aware* that some options may be perceived as prominent by people from the same culture. In fact, if individuals think that their assigned partners are culturally alike, then they expect them to view the problem in a similar way. So, if individuals have reason to believe that an option easily comes to mind to people with the same linguistic background, then they are more likely to select that option in coordination games.

[21] For example, consider games with identical payoff structures but different strategy labels, such as {"cooperate", "defect"} versus {"out", "in"}. For a guilt-aversion account of the role of labels in social dilemmas, see Dufwenberg, Gächter, and Hennig-Schmidt (2011); see also Ellingsen, Johannesson, Mollerstrom, and Munkhammar (2012).

APPENDIX A

Additional figures and tables

<u>Study 1</u>

Figure 1A shows frequency distributions of individual-level choices in each game (for the list of labels in each of the 1-10 games, in each Version A and B, see Figure 2 in the body of the paper). Note that – for visual clarity – here we have arranged the order of the options in each game, in such a way that our predicted choice (i.e., the label with the highest NGRAM) is always shown at the bottom of the bar.

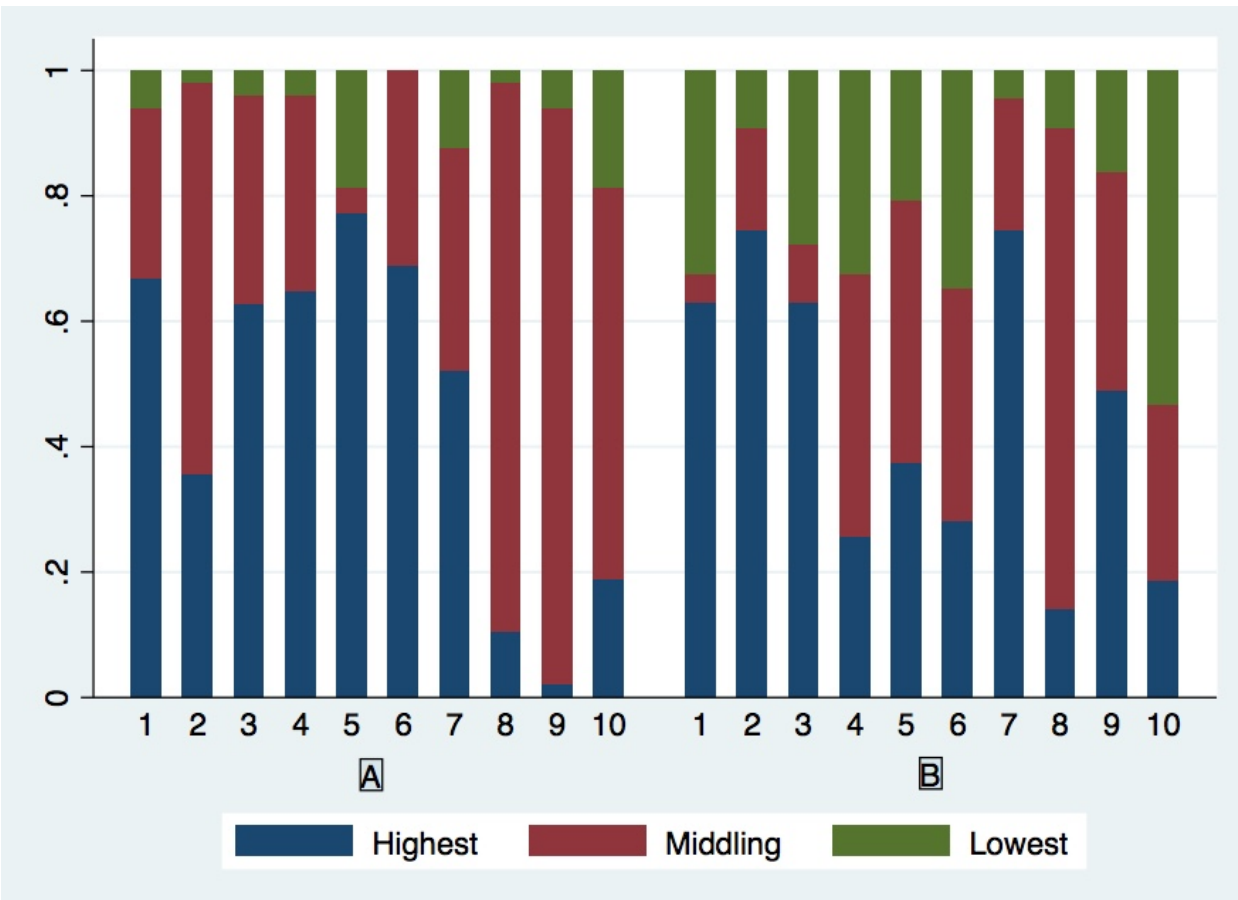


Figure 1A - Frequency distributions of individual-level choices in each of the 1-10 games in Version A (left panel) and Version B (right panel). The blue, maroon and green bars indicate the frequency with which the option with the relatively highest, middling and lowest NGRAM, respectively, was chosen (note that the option with the highest NGRAM is always shown at the bottom of the bar).

Study 2

Figure 2A shows frequency distributions of individual-level choices in each of the 1-10 games and for each of the 1-4 conditions (for the labels, see the left panel of Figure 2 in the body of the paper). Note that here we have arranged the order of the options in each game, in such a way that the most frequently-mentioned label (i.e., the one with the highest NGRAM) is always shown at the bottom of the bar.

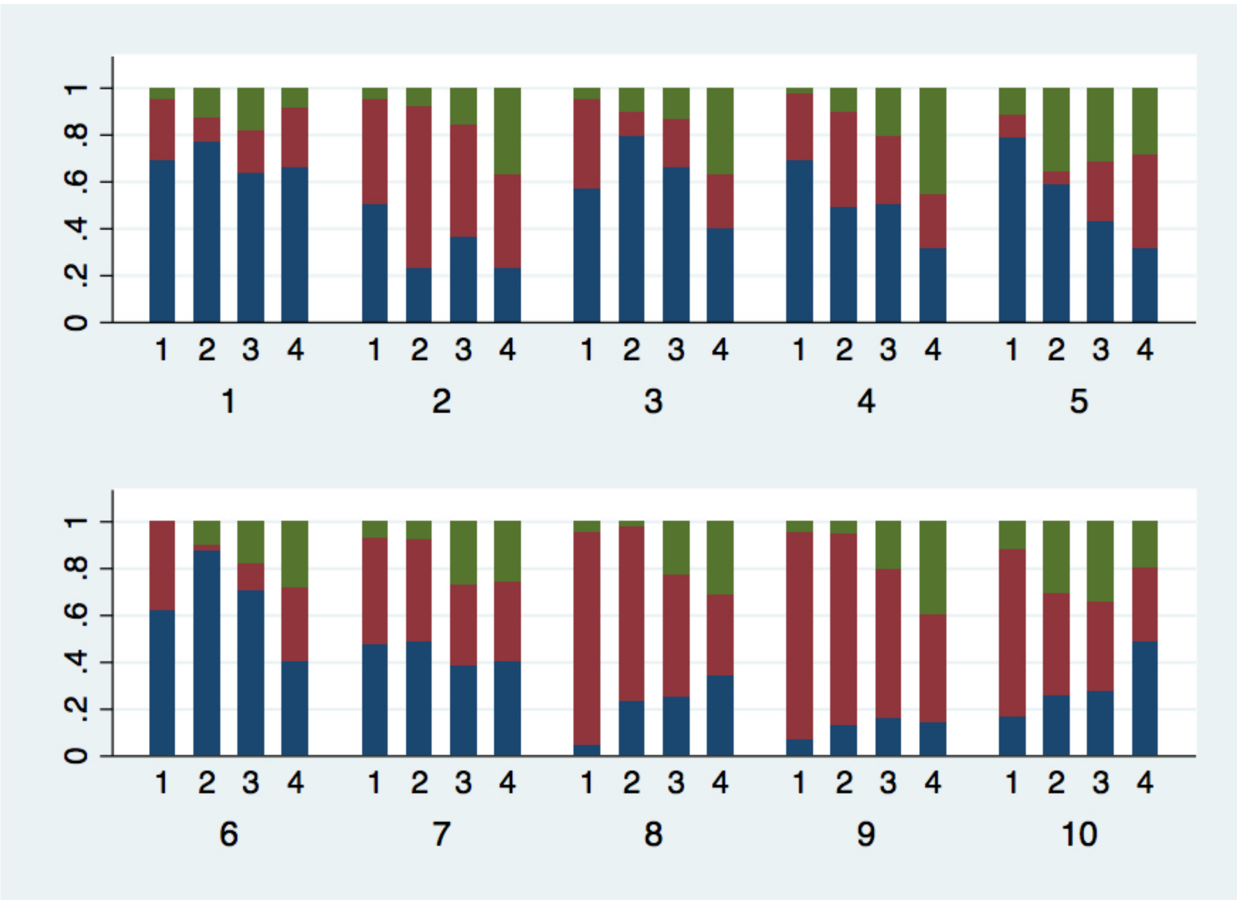


Figure 2A - Frequency distributions of individual-level choices in each of the *1-10 games* and for each of the *1-4 conditions*. NOTE: Conditions 1-4 refer to “Coordinate”, “Pick”, “Seek”, and “Hide”, respectively. The blue, maroon and green bars respectively indicate the frequency with which the option with the relatively highest, middling and lowest NGRAM was chosen (note that the option with the highest NGRAM is always shown at the bottom of the bar).

In what follows we present a few conditional logistic models. Before discussing the results, we note that conditional logistic regressions differ from regular logistic analysis in that the data are grouped (so that all observations in a group have a common characteristic), and hence the likelihood is computed relative to each group. More precisely, below we present alternative-specific conditional logistic models (asclogit—McFadden, 1973), where the likelihood of a subject's choice may depend on: *(i)* "alternative-specific" regressors, and *(ii)* "case-specific" regressors. Each model below features a central-upper section (presenting alternative-specific regressors) and a bifurcated-lower section (presenting case-specific regressors). Let's start from the latter section.

Case-specific regressors refer to characteristics that are common to all alternatives in a group (i.e., triplet of labels), such as the experimental condition "C" to which one is assigned. These coefficients are interpreted as parameters of an ordinary multinomial logit model, against the base category (which is specified directly above the bifurcation).

Alternative-specific regressors refer to attributes that are specific to an option – i.e., label – as opposed to being specific to the whole triplet of options. Alternative-specific coefficients measure the impact of a change in an option's attribute (on the likelihood of choosing that option). If a regressor has a positive coefficient, then an increase in that attribute (e.g., an increase in the relative word frequency of a label) implies that the corresponding strategy is selected more often than the other two strategies in the triplet, whichever they are. Thus, the output of each asclogit model presents a unique coefficient for each alternative-specific regressor, without an explicit base category (Cameron and Trivedi, 2010, pp. 503-511).

All the models below involve the following predictors: a continuous variable "WF" measuring the relative word frequency of a label (defined as the *ratio of the n-gram frequency of the label to the n-gram frequencies of all the labels in the triplet*); a dummy "C" indicating whether an option is shown to a participant in the alternative condition (as specified in the table notes); an interaction variable "WFC".

To perform a pairwise comparison of choice behavior across conditions, we begin by discussing model [1] – in the left panel of Table 1A – which uses the full set of option triplets (i.e., ten games per subject) in the *Pick* and *Coordinate* conditions. There, the strong significance of *WF* confirms a positive effect of word frequency on choice behavior: this means that *the higher the relative n-gram frequency of a label, the more likely it is for the associated strategy to be selected, regardless of any other characteristics* (e.g., experimental condition or label's position on the screen). We further note that *WFC* is non-significant: this means that there is no difference in the relative impact of word frequency between Pickers and Coordinators (we stress that regressor *WFC* is alternative specific, as it captures how the impact of the n-gram frequency of a particular label varies with the experimental condition).

A few more comments are in order. The two *constant* terms of model [1] (i.e., -.357 and -.098) respectively indicate the desirability of the options located at the middle and bottom of the list on the screen (against the top option), due to unmeasured attributes of the alternatives: they are negative (though one is non-significant), reflecting the greater desirability of the options located at the top of the list. Lastly, the main-effects coefficients for the regressor *C* (i.e., -.948 and -.853) indicate the impact of a change in experimental condition on the desirability of the middle and bottom options (against the top option): they are both negative, revealing that middle and bottom options are selected less often by Coordinators than Pickers.

Choice of option	**[1] *Pick & Coordinate***		**[2] *Coordinate & Seek***	
Relative word frequency ("WF")	.637*** (.161)		.707*** (.111)	
Relative word frequency * *Condition* ("WFC")	-.050 (.104)		-.119** (.049)	
	Center vs. Top	**Bottom vs. Top**	**Center vs. Top**	**Bottom vs. Top**
Condition ("C")	-.948*** (.212)	-.853*** (.212)	1.157*** (.227)	.896*** (.214)
Constant	-.357*** (.115)	-.098 (.110)	-1.305*** (.177)	-.951*** (.181)
Log pseudolikelihood	-746.093		-824.402	
Obs.	810		860	

Table 1A - Alternative-specific conditional logistic regression coefficients, with clustered standard errors in brackets (*, **, and *** indicate $p<0.10$, $p<0.05$ and $p<0.01$, respectively, for the relevant Z Statistic, two-tailed tests).
Model [1]: regressor *C* takes on value 0 if a subject is assigned to *Pick*; it takes on value 1 if a subject is assigned to *Coordinate*.
Model [2]: regressor *C* takes on value 0 if a subject is assigned to *Coordinate*; it takes on value 1 if a subject is assigned to *Seek*.

We move on to perform a few more pairwise comparisons: Model [2], in the right panel of Table 1A above, uses the full set of triplets from the *Coordinate* and *Seek* conditions. Model [3], in the left panel of Table 2A below, uses the full set of triplets from the *Pick* and *Hide* conditions. Finally Model [4] below uses the full set of triplets from the *Seek* and *Hide* conditions. In short, we note that *WF* is positive and significant in all models. We also note that *WFC* is negative and significant in models [2], [3], [4]: this means that the most frequently-mentioned label is selected less often when moving from *Coordinate* to *Seek* (model [2]), from *Pick* to *Hide* (model [3]), and from *Seek* to *Hide* (model [4]), thereby confirming our hypotheses.

Choice of option	**[3]** ***Pick & Hide***		**[4]** ***Seek & Hide***	
Relative word frequency ("WF")	.940*** (.171)		.991*** (.350)	
Relative word frequency * *Condition* ("WFC")	-.201*** (.053)		-.214** (.100)	
	Center vs. Top	**Bottom vs. Top**	**Center vs. Top**	**Bottom vs. Top**
Condition ("C")	.535*** (.174)	.169 (.168)	.325* (.192)	.126 (.172)
Constant	-.357*** (.115)	-.098 (.110)	-.148 (.141)	-.054 (.115)
Log pseudolikelihood	-767.819		-846.129	
Obs.	740		790	

Table 2A - Alternative-specific conditional logistic regression coefficients, with clustered standard errors in brackets (*, **, and *** indicate $p<0.10$, $p<0.05$ and $p<0.01$, respectively, for the relevant Z Statistic, two-tailed tests).
Model [3]: regressor *C* takes on value 0 if a subject is assigned to *Pick*; it takes on value 1 if a subject is assigned to *Hide*.
Model [4]: regressor *C* takes on value 0 if a subject is assigned to *Seek*; it takes on value 1 if a subject is assigned to *Hide*.

Study 3

Figures 3A and 4A below respectively show frequency distributions of individual-level choices (in each of the 1-10 games and for each of the 1-3 conditions) in the US and UK samples (for the labels, see Figure 4 in the body of the paper). Note that – for visual clarity – we have arranged the order of the options in each game in such a way that the option with the highest NGRAM *in the relevant vocabulary* is shown at the bottom of the bar.

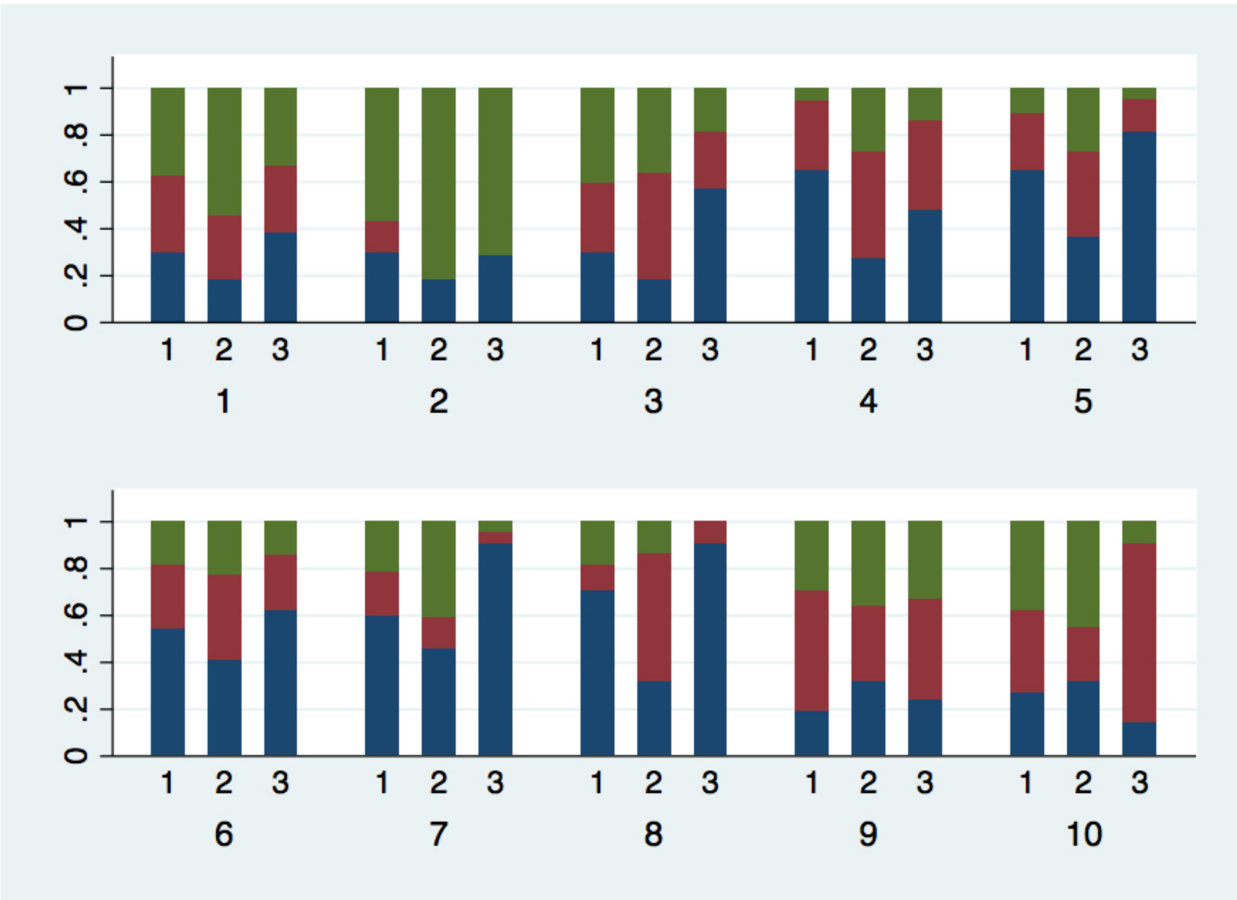


Figure 3A - ***US sample***. Frequency distributions of individual-level choices for each of the *1-10 games* and for each of the *1-3 information conditions*. Conditions 1-3 refer to: “NO-k”, “k-UK”, and “k-US”, respectively. The blue, maroon and green bars respectively indicate the frequency with which the option with the relatively highest, middling and lowest American-English NGRAM was chosen (i.e., the label with the highest American-English NGRAM is always shown at the bottom of the bar).

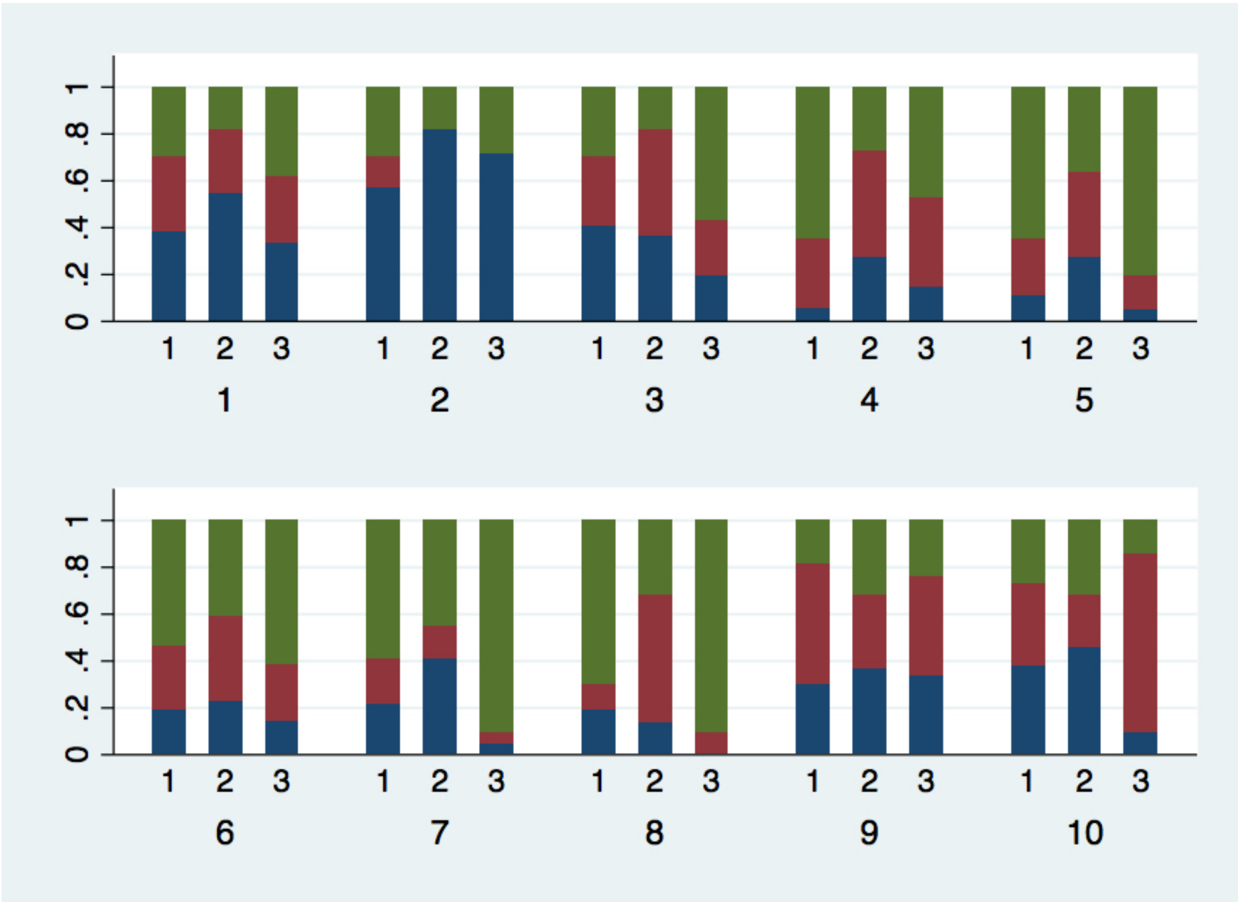


Figure 4A - *UK sample*. Frequency distributions of individual-level choices for each of the *1-10 games* and for each of the *1-3 information conditions*. Conditions 1-3 refer to: "NO-k", "k-UK", and "k-US", respectively. The blue, maroon and green bars respectively indicate the frequency with which the option with the relatively highest, middling and lowest British-English NGRAM was chosen (i.e., the label with the highest British-English NGRAM is always shown at the bottom of the bar).

In what follows we present alternative-specific conditional logistic (asclogit) regressions. (Please refer to p. 32 above for a general commentary on the asclogit model.) Model [5] of Table 3A below refers to the *US sample* and uses the full set of option triplets (i.e., ten games per subject) from the *k-UK* and *k-US* conditions. By contrast, model [6] of Table 3A refers to the *UK sample* and uses the full set of option triplets from the very same conditions. Both regressions involve the following predictors: a continuous variable "WF" measuring the relative word frequency of a strategy-label in the relevant vocabulary (i.e., *American-English* NGRAM in model [5]; *British-English* NGRAM in model [6]); a dummy "C", which takes on value 0 (1) if a subject is informed that her counterpart resides in the same (different) country; an interaction variable "WFC".

Choice of option	[5] *k-UK & k-US*		[6] *k-UK & k-US*	
Relative word frequency ("WF")	4.479*** (.747)		1.841** (.861)	
Relative word frequency * *Condition* ("WFC")	-4.316*** (1.471)		-4.218*** (1.436)	
	Center vs. Top	**Bottom vs. Top**	**Center vs. Top**	**Bottom vs. Top**
Condition ("C")	.193 (.550)	.478 (.508)	-.261 (.412)	-.750* (.437)
Constant	-.491 (.395)	-1.078*** (.316)	-.233 (.313)	-.496 (.329)
Log pseudolikelihood	-203.014		-220.462	
Obs.	210		220	

Table 3A - Alternative-specific conditional logistic regression coefficients, with clustered standard errors in brackets (*, **, and *** indicate $p<0.10$, $p<0.05$ and $p<0.01$, respectively, for the relevant Z Statistic, two-tailed tests).
Model [5]: ***US sample***. *WF* refers to the American-English NGRAM. Regressor *C* takes on value 0 if a subject is assigned to *k-US*; it takes on value 1 if a subject is assigned to *k-UK*.
Model [6]: ***UK sample***. *WF* refers to the British-English NGRAM. Regressor *C* takes on value 0 if a subject is assigned to *k-UK*; it takes on value 1 if a subject is assigned to *k-US*.

Note that *WF* is positive and significant in both regressions: this means that the higher the relative n-gram frequency of a label (in the relevant vocabulary), the more likely it is for the associated strategy to be selected, regardless of any other characteristics. Further, note that *WFC* is negative and significant in both regressions: this means that if one is informed that the assigned partner resides in a different country, one will less likely choose the strategy associated with the most frequently-mentioned label (in one's own vocabulary).

APPENDIX B

Experimental instructions and screen shots

<u>Study 1</u>

INSTRUCTIONS

Welcome. In this study we are interested in how people coordinate. Please read the following instructions carefully. Note that you may earn a bonus payment in the amount of £0.10 per question, as follows. Specifically, the bonus payment will be determined by your own choices and those made by some other participant, according to the rules described below.

In this experiment each participant (a *Prolific* worker) will be given the same choice tasks, and each participant will be assigned a partner at random (another *Prolific* worker). Note that you will not know the identity of your partner nor will you be able to communicate with him/her.

Both you and your partner will be given 10 multiple-choice questions, with three options in each. If you both select the same option in a given question you will have "coordinated." If you select different options in a given question you will have "miscoordinated." Your goal is to select the same option as your partner (that is, to coordinate with your partner) on as many questions as possible. Recall that you will not know the identity of your partner nor will you be able to communicate with him/her.

Hence, consider a question that asks you to select an option among three, for example, *green*, *blue* and *yellow*. If both you and your partner select the same option (i.e. you both select green, or you both select blue, or you both select yellow) then you will have coordinated. If you select different options (e.g. you select green and your partner selects blue) then you will have miscoordinated.

In short, you will receive a bonus payment of £0.10 per coordinated answer. So - given that this study comprises 10 questions - the total range of bonus payments you can expect to receive is [£0, £1].

>>

Below is a screen shot of a typical experimental task.

Please choose one option. Each of you and your partner will receive £0.10 if you both choose the same option, £0 otherwise.

- butter
- peppermint
- cocoa

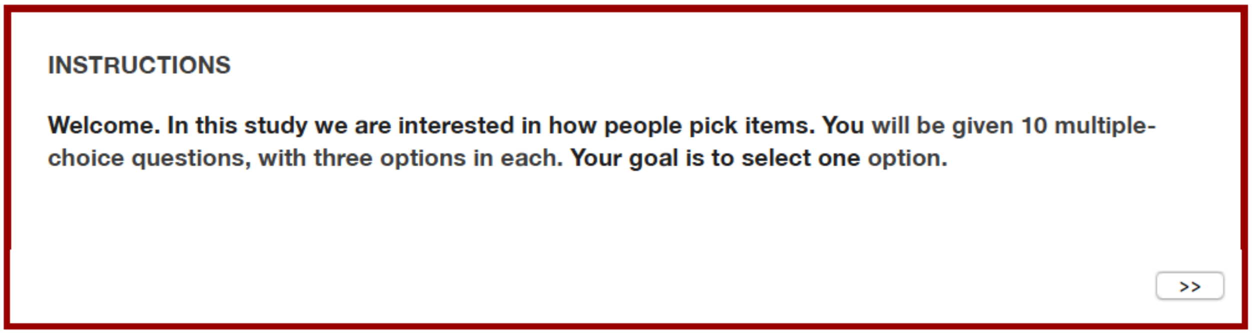
>>

Study 2

Coordinate

Same instructions as in Study 1.

Pick



INSTRUCTIONS

Welcome. In this study we are interested in how people pick items. You will be given 10 multiple-choice questions, with three options in each. **Your goal is to select one** option.

>>

Below is a screen shot of a typical experimental task.

Please pick one option.

- Nepal
- Uzbekistan
- Thailand

>>

Hide-and-Seek

INSTRUCTIONS

Welcome. In this study we are interested in how people choose among options. Please read the following instructions carefully. Note that you may earn a bonus payment in the amount of £0.10 per question, as follows. Specifically, the bonus payment will be determined by your own choices and those made by some other participant, according to the rules described below.

In this experiment each participant (a *Prolific* worker) will be randomly assigned one of two roles (either "Hider" or "Seeker") and matched with a partner of a different role (another *Prolific* worker). Note that you will not know the identity of your partner nor will you be able to communicate with him/her.

Both you and your partner will be given 10 multiple-choice questions, with three options in each. In each question (whether you are assigned the role of Hider or Seeker) you will have to select one of the three options.

Specifically, if you are assigned the role of *Seeker*, your objective will be to select the *same* option as your partner.

Instead, if you are assigned the role of *Hider*, your objective will be to select a *different* option from your partner.

(Recall that you will not be able to communicate with him/her.)

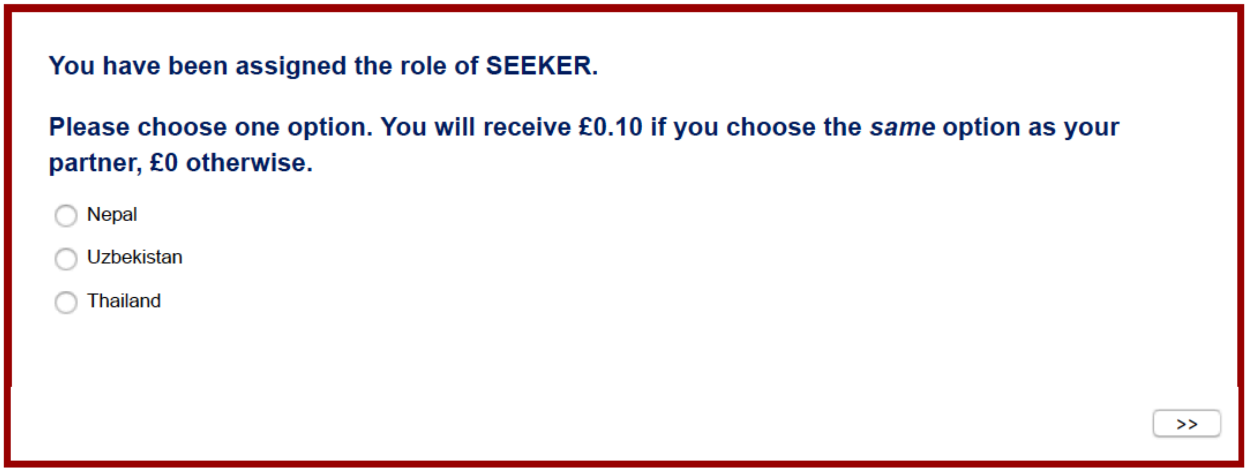
For example, consider a question that asks you to select an option among, say, *green*, *blue* and *yellow*. If both you and your partner select the same option (i.e. you both select green, or you both select blue, or you both select yellow), then the Seeker will receive a bonus payment of £0.10 while the Hider will receive nothing.

On the other hand, if you and your partner select a different option (e.g. you select green and instead your partner selects blue or yellow), then the Hider will receive a bonus payment of £0.10 while the Seeker will receive nothing.

In summary, you may receive a bonus payment of £0.10 per answer depending on your role and the choices of you and your partner. Hence - given that this study comprises 10 questions - the total range of bonus payments you can expect to receive is [£0, £1].

>>

Below is a screen shot of a typical experimental task (***Seeker***).



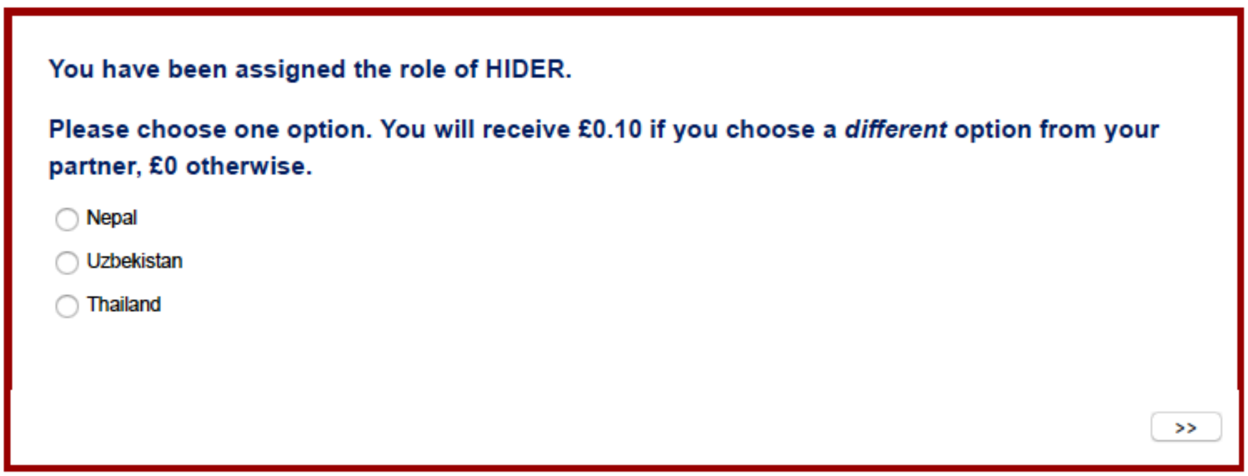
You have been assigned the role of SEEKER.

Please choose one option. You will receive £0.10 if you choose the *same* option as your partner, £0 otherwise.

- Nepal
- Uzbekistan
- Thailand

>>

Below is a screen shot of a typical experimental task (***Hider***).



You have been assigned the role of HIDER.

Please choose one option. You will receive £0.10 if you choose a *different* option from your partner, £0 otherwise.

- Nepal
- Uzbekistan
- Thailand

>>

Study 3

INSTRUCTIONS

Welcome. In this study we are interested in how people coordinate. Please read the following instructions carefully. Note that you may earn a bonus payment in the amount of £0.10 per question, as follows. Specifically, the bonus payment will be determined by your own choices and those made by some other participant, according to the rules described below.

In this experiment each participant (a *Prolific* worker) will be given the same choice tasks, and each participant will be assigned a partner at random (another *Prolific* worker). Note that you will not know the identity of your partner nor will you be able to communicate with him/her.

Both you and your partner will be given 10 multiple-choice questions, with three options in each. If you both select the same option in a given question you will have "coordinated." If you select different options in a given question you will have "miscoordinated." Your goal is to select the same option as your partner (that is, to coordinate with your partner) on as many questions as possible. Recall that you will not know the identity of your partner nor will you be able to communicate with him/her.

Hence, consider a question that asks you to select an option among three, for example, *green*, *blue* and *yellow*. If both you and your partner select the same option (i.e. you both select green, or you both select blue, or you both select yellow) then you will have coordinated. If you select different options (e.g. you select green and your partner selects blue) then you will have miscoordinated.

In short, you will receive a bonus payment of £0.10 per coordinated answer. So - given that this study comprises 10 questions - the total range of bonus payments you can expect to receive is [£0, £1].

>>

Below is a screen shot of a typical experimental task (***No-k***).

Please choose one option. Each of you and your partner receive £0.10 if you both choose the same option, £0 otherwise.

- peach
- pineapple
- pear

>>

Below is a screen shot of a typical experimental task (***k-UK)***.

Please choose one option. Each of you and your partner receive £0.10 if you both choose the same option, £0 otherwise.

Your partner is a Prolific worker who resides in the UK. Your partner may or may not know where you reside.

- peach
- pineapple
- pear

>>

Below is a screen shot of a typical experimental task (***k-US)***.

Please choose one option. Each of you and your partner receive £0.10 if you both choose the same option, £0 otherwise.

Your partner is a Prolific worker who resides in the US. Your partner may or may not know where you reside.

- peach
- pineapple
- pear

>>

References

Alós-Ferrer, Carlos and Christoph Kuzmics. 2013. "Hidden Symmetries and Focal Points" *Journal of Economic Theory*, 148(1): 226-258.

Anderson, John R. and Lael J. Schooler. 1991. "Reflections of the Environment in Memory" *Psychological Science*, 2(6): 396-408.

Arad, Ayala and Ariel Rubinstein. 2012. "The 11-20 Money Request Game: A Level-k Reasoning Study" *American Economic Review*, 102(7): 3561-73.

Bacharach, Michael. 1993. "Variable Universe Games" in *Frontiers of Game Theory*, ed. K. Binmore et al. Cambridge, MA: MIT Press.

Bacharach, Michael. 2006. Beyond Individual Choice: Teams and Frames in Game Theory, ed. N. Gold and R. Sugden. Princeton University Press.

Bacharach, Michael and Michele Bernasconi. 1997. "The Variable Frame Theory of Focal Points: An Experimental Study" *Games and Economic Behavior*, 19(1): 1-45.

Bacharach, Michael and Dale O. Stahl. 2000. "Variable-Frame Level-n Theory" *Games and Economic Behavior*, 32(2): 220-246.

Balota, David A. and Daniel H. Spieler. 1999. "Word Frequency, Repetition, and Lexicality Effects in Word Recognition Tasks: Beyond Measures of Central Tendency" *Journal of Experimental Psychology: General*, 128(1): 32-55.

Bardsley, Nicholas, Judith Mehta, Chris Starmer and Robert Sugden. 2010. "Explaining Focal Points: Cognitive Hierarchy Theory versus Team Reasoning" *Economic Journal*, 120(543): 40-79.

Bardsley, Nicholas and Aljaž Ule. 2017. "Focal Points Revisited: Team Reasoning, the Principle of Insufficient Reason and Cognitive Hierarchy Theory" *Journal of Economic Behavior & Organization*, 133: 74-86.

Blume, Andreas. 2000. "Coordination and Learning with a Partial Language" *Journal of Economic Theory*, 95(1): 1-36.

Camerer, Colin F., Teck-Hua Ho and Juin-Kuan Chong. 2004. "A Cognitive Hierarchy Model of Games" *Quarterly Journal of Economics*, 119(3): 861-98.

Camerer, Colin F. and Marc Knez. 1997. "Coordination in Organizations: A Game-Theoretic Perspective" in *Organizational Decision Making*, ed. Z. Shapira. Cambridge University Press.

Cameron, A. Colin and Pravin K. Trivedi. 2010. *Microeconometrics Using Stata, Revised Edition*. College Station, TX: Stata Press.

Casajus, André. 2000. "Focal Points in Framed Strategic Forms" *Games and Economic Behavior*, 32(2): 263-291.

Charness, Gary, Uri Gneezy and Michael A. Kuhn. 2013. "Experimental Methods: Extra-Laboratory Experiments—Extending the Reach of Experimental Economics" *Journal of Economic Behavior & Organization*, 91: 93-100.

Cooper, Russell W., Douglas V. DeJong, Robert Forsythe and Thomas W. Ross. 1990. "Selection Criteria in Coordination Games: Some Experimental Results" *American Economic Review*, 80(1): 218-233.

Costa-Gomes, Miguel A. and Vincent P. Crawford. 2006. "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study" *American Economic Review*, 96(5): 1737-68.

Costa-Gomes, Miguel A., Vincent P. Crawford and Bruno Broseta. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study" *Econometrica*, 69(5): 1193-1235.

Crawford, Vincent P., Miguel A. Costa-Gomes and Nagore Iriberri. 2013. "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications" *Journal of Economic Literature*, 51(1): 5-62.

Crawford, Vincent P., Uri Gneezy and Yuval Rottenstreich. 2008. "The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures" *American Economic Review*, 98(4): 1443-58.

Crawford, Vincent P. and Hans Haller. 1990. "Learning How to Cooperate: Optimal Play in Repeated Coordination Games" *Econometrica*, 58(3): 571-595.

Crawford, Vincent P. and Nagore Iriberri. 2007. "Fatal Attraction: Salience, Naivete, and Sophistication in Experimental Hide-and-Seek Games" *American Economic Review*, 97(5): 1731-1750.

Cubitt, Robin P. and Robert Sugden. 2003. "Common Knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory" *Economics & Philosophy*, 19(2): 175-210.

Dougherty, Michael R., Ana M. Franco-Watkins and Rick Thomas. 2008. "Psychological Plausibility of the Theory of Probabilistic Mental Models and the Fast and Frugal Heuristics" *Psychological Review*, 115(1): 199-211.

Dufwenberg, Martin, Simon Gächter and Heike Hennig-Schmidt. 2011. "The Framing of Games and the Psychology of Play" *Games and Economic Behavior*, 73(2): 459-478.

Ellingsen, Tore, Magnus Johannesson, Johanna Mollerstrom and Sara Munkhammar. 2012. "Social Framing Effects: Preferences or Beliefs?" *Games and Economic Behavior*, 76(1): 117-130.

Faillo, Marco, Alessandra Smerilli and Robert Sugden. 2017. "Bounded Best-Response and Collective-Optimality Reasoning in Coordination Games" *Journal of Economic Behavior & Organization*, 140: 317-335.

Gigerenzer, Gerd and Daniel G. Goldstein. 1996. "Reasoning the Fast and Frugal Way: Models of Bounded Rationality" *Psychological Review*, 103(4): 650-669.

Goldstein, Daniel G. and Gerd Gigerenzer. 2002. "Models of Ecological Rationality: The Recognition Heuristic" *Psychological Review*, 109(1): 75-90.

Hargreaves Heap, Shaun, David Rojo Arjona and Robert Sugden. 2014. "How Portable Is Level-0 Behavior? A Test of Level-k Theory in Games with Non-Neutral Frames" *Econometrica*, 82(3): 1133-1151.

Hargreaves Heap, Shaun, David Rojo Arjona and Robert Sugden. 2017. "Coordination When There Are Restricted and Unrestricted Options" *Theory and Decision*, 83(1): 107-129.

Harsanyi, John C. and Reinhard Selten. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

Hedden, Trey and Jun Zhang. 2002. "What Do You Think I Think You Think?: Strategic Reasoning in Matrix Games" *Cognition*, 85(1): 1-36.

Hertwig, Ralph, Stefan M. Herzog, Lael J. Schooler and Torsten Reimer. 2008. "Fluency Heuristic: A Model of How the Mind Exploits a By-Product of Information Retrieval" *Journal of Experimental Psychology: Learning, memory, and cognition*, 34(5): 1191-1206.

Hertwig, Ralph, Thorsten Pachur and Stephanie Kurzenhäuser. 2005. "Judgments of Risk Frequencies: Tests of Possible Cognitive Mechanisms" *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(4): 621-642.

Isoni, Andrea, Anders Poulsen, Robert Sugden and Kei Tsutsui. 2013. "Focal Points in Tacit Bargaining Problems: Experimental Evidence" *European Economic Review*, 59: 167-188.

Janssen, Maarten C. W. 2001. "Rationalizing Focal Points" *Theory and Decision*, 50(2): 119-148.

Jescheniak, Jörg D. and Willem J. M. Levelt. 1994. "Word Frequency Effects in Speech Production: Retrieval of Syntactic Information and of Phonological Form" *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4): 824-843.

Kapferer, Jean-Noel. 2012. The New Strategic Brand Management: Advanced Insights and Strategic Thinking. London: Kogan Page.

Lewis, David K. 1969. Convention: A Philosophical Study. Cambridge, MA: Harvard U. Press.

Lichtenstein, Sarah, Paul Slovic, Baruch Fischhoff, Mark Layman and Barbara Combs. 1978. "Judged Frequency of Lethal Events" *Journal of Experimental Psychology: Human Learning and Memory*, 4(6): 551-578.

McFadden, Daniel. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior" in *Frontiers in Econometrics*, ed. P. Zarembka. New York: Academic Press.

Mehta, Judith, Chris Starmer and Robert Sugden. 1994. "The Nature of Salience: An Experimental Investigation of Pure Coordination Games" *American Economic Review*, 84(3): 658-73.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, the Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak and Erez Lieberman Aiden. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books" *Science*, 331(6014): 176-182.

Morris, Stephen and Hyun Song Shin. 1997. "Approximate Common Knowledge and Co-ordination: Recent Lessons from Game Theory" *Journal of Logic, Language and Information*, 6(2): 171-190.

Morrison, Catriona M. and Andrew W. Ellis. 1995. "Roles of Word Frequency and Age of Acquisition in Word Naming and Lexical Decision" *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1): 116-133.

Nagel, Rosemarie. 1995. "Unraveling in Guessing Games: An Experimental Study" *American Economic Review*, 85(5): 1313-26.

Penczynski, Stefan P. 2016. "Strategic Thinking: The Influence of the Game" *Journal of Economic Behavior & Organization*, 128: 72-84.

Rubinstein, Ariel and Yuval Salant. 2016. "'Isn't Everyone like Me?': On the Presence of Self-Similarity in Strategic Interactions" *Judgment and Decision Making*, 11(2): 168-173.

Rubinstein, Ariel, Amos Tversky and Dana Heller. 1997. "Naive Strategies in Competitive Games" in *Understanding Strategic Interaction*, ed. W. Albers et al. Berlin: Springer-Verlag.

Schelling, Thomas C. 1960. The Strategy of Conflict. Cambridge, MA: Harvard U. Press.

Schooler, Lael J. and Ralph Hertwig. 2005. "How Forgetting Aids Heuristic Inference" *Psychological Review*, 112(3): 610-628.

Schwarz, Norbert, Herbert Bless, Fritz Strack, Gisela Klumpp, Helga Rittenauer-Schatka and Annette Simons. 1991. "Ease of Retrieval as Information: Another Look at the Availability Heuristic" *Journal of Personality and Social Psychology*, 61(2): 195-202.

Sebanz, Natalie, Günther Knoblich and Wolfgang Prinz. 2003. "Representing Others' Actions: Just Like One's Own?" *Cognition*, 88(3): B11-B21.

Seidenberg, Mark S. and James L. McClelland. 1989. "A Distributed, Developmental Model of Word Recognition and Naming" *Psychological Review*, 96(4): 523-568.

Stahl, Dale O. and Paul W. Wilson. 1995. "On Players' Models of Other Players: Theory and Experimental Evidence" *Games and Economic Behavior*, 10(1): 218-54.

Sugden, Robert. 1995. "A Theory of Focal Points" *Economic Journal*, 105(430): 533-550.

Sugden, Robert and Ignacio E. Zamarrón. 2006. "Finding the Key: The Riddle of Focal Points" *Journal of Economic Psychology*, 27(5): 609-621.

Tversky, Amos and Daniel Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability" *Cognitive Psychology*, 5(2): 207-232.

Tversky, Amos and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases" *Science*, 185(4157): 1124-1131.

Van Huyck, John B., Raymond C. Battalio and Richard O. Beil. 1990. "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure" *American Economic Review*, 80(1): 234-248.

Weber, Roberto A. and Colin F. Camerer. 2003. "Cultural Conflict and Merger Failure: An Experimental Approach" *Management Science*, 49(4): 400-415.